REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503

1. Agency Use Only (Leave blank)	2. Report Date	3. Report Type and Period Covered (i.e., annual 1 Jun 00 - 31 May 01)
	May 15, 2001	Final Report (3/1/97 - 3/1/01)

4. Title and Subtitle
Selection and Computational Potential of Gene Control Elements and Their Circuitry

5. Award Number
N00014-97-1-0364 (G)

6. Author(s)
Michael A. Savageau

7. Performing Organization Name (Include Name, City, State, Zip Code and Email for Principal Investigator)
Department of Microbiology and Immunology
5641 Medical Science II, Box 0620
University of Michigan Medical School
Ann Arbor, MI 48109-0620
E-Mail: savageau@umich.edu

8. Performing Organization Report Number (Leave Blank)

9. Sponsoring/Monitoring Agency Name and Address
Office of Naval Research
ONR 342CN
Ballston Centre Tower One
800 North Quincy Street
Arlington, VA 22217-5660

10. Sponsoring/Monitoring Agency Report Number (Leave Blank)

11. Supplementary Notes (i.e., report contains color photos, report contains appendix in non-print form, etc.)
N/A

12a. Distribution/Availability Statement (check one)
[X] Approved for public release; distribution unlimited
[ ] Distribution limited to U.S. Government agencies only - report contains proprietary information

12b. Distribution Code
(Leave Blank)

13. Abstract ***(Maximum 200 Words)*** *(abstract should contain no proprietary or confidential information)*

The physical basis for biological complexity is context-dependent expression of the organism's genome. The context is provided by the life cycle of the organism; the molecular mechanisms of gene regulation interpret that context. We seek to develop a theoretical framework for quantitatively relating the integrated behavior of gene circuits to their underlying molecular determinants, and, by applying this theory to specific classes of gene circuits, we hope to discover the basic principles that govern their design by natural selection. To the best of our knowledge, there has been no concerted effort to understand the role of gene circuitry as a robust computational device. Under what conditions will gene control elements and their circuitry be maintained in the face of mutational entropy? What is the computational potential of such robust circuits, and can one design selection strategies to direct their evolution? These related questions represent the principal objectives of this work. We will first analyze known molecular modes of gene control and the spectrum of computations that they are capable of performing. Second, we will determine the selective pressures that in the presence of mutation lead to the emergence and maintenance of particular computational circuits involving these elements. Finally, we will use all this information in an attempt to design specific computational solutions through a process of directed evolution.

14. Subject Terms (keywords previously assigned to proposal abstract or terms which apply to this award)
Control of gene expression, Computational potential, Life cycle, Design principles, Robust circuitry

15. Number of Pages (count all pages including appendices)

16. Price Code

17. Security Classification of Report	18. Security Classification of this Page	19. Security Classification of Abstract	20. Limitation of Abstract
Unclassified	Unclassified	Unclassified	Unlimited

FINAL TECHNICAL REPORT

Long-term Research Objective:

The physical basis for biological complexity is context-dependent expression of the organism's genome. The context is provided by the life cycle of the organism; the molecular mechanisms of gene regulation interpret that context. We seek to develop a theoretical framework for quantitatively relating the integrated behavior of gene circuits to their underlying molecular determinants, and, by applying this theory to specific classes of gene circuits, we hope to discover the basic principles that govern their design by natural selection.

S & T Objectives:

To the best of our knowledge, there has been no concerted effort to understand the role of gene circuitry as a robust computational device. Under what conditions will gene control elements and their circuitry be maintained in the face of mutational entropy? What is the computational potential of such robust circuits, and can one design selection strategies to direct their evolution? These related questions represent the principal objectives of this work.

Approach:

We will first analyze known molecular modes of gene control and the spectrum of computations that they are capable of performing. Second, we will determine the selective pressures that in the presence of mutation lead to the emergence and maintenance of particular computational circuits involving these elements. Finally, we will use all this information in an attempt to design specific computational solutions through a process of directed evolution.

S & T Completed:

Analysis of Alternative Molecular Modes of Gene Control And of The Selective Pressures That Lead to The Emergence And Maintenance of Particular Computational Circuits

We have refined our development of the quantitative implications of demand theory. A summary of the preliminary results that were obtained in the first year of the grant appeared as a book chapter [1]. The detailed development of the theory and applications with refined parameter estimates were presented in two *Genetics* papers. In the first paper [2], a theory is developed that ties together a number of important variables, including growth rates, mutation rates, minimum and maximum demands for gene expression, and minimum and maximum durations for the life cycle of the organism. Applications of the theory are provided in the second paper [3], where regulation of the lactose and maltose operons of *Escherichia coli* is analyzed and the results are compared with independent experimental data.

The quantitative development of demand theory not only confirms and quantifies the previous qualitative predictions, but it also identifies critical factors and reveals new relationships [2].

20010604 129

The recursive equations that characterize the population dynamics of mutant and wild-type organisms allow one to predict the time course for selection. The form of these equations also allows one to predict that the response time for selection is independent of the ON/OFF cycle time for the gene C, whereas it is strongly dependent upon the demand for gene expression D. The steady-state solution of the recursive equations provides estimates for the extent of selection. A threshold for selection is determined by the relationship between cycle time C and demand D that results when the extent of selection is set equal to the criterion for selection.

The thresholds for selection in the C *vs.* D plot define regions within which selection of the positive or negative mode of regulation is realizable. Their intersection defines a maximum value for the cycle time C_{max}, and their asymptotes define minimum D_{min} and maximum D_{max} values of the demand for gene expression. These regions also exhibit an inherent asymmetry that favors selection of the positive mode.

In summary, the quantitative development of demand theory reveals unexpected relationships between the demand for gene expression D and the average ON/OFF cycle time for the gene C, which is a manifestation of the organism's life cycle. The demand theory of gene regulation can be extended within the framework presented here to include organisms with life cycles that are more complex than the two phases illustrated in this paper and regulatory systems that are more complex than a single mechanism of gene control.

The application of demand theory to the lactose and maltose operons of *E. coli* provides an opportunity to test a number of the theory's quantitative implications [3].

With the parameter values that represent the lactose and maltose operons, the time required to reach full selection is independent of cycle time but decreases until a minimum is reached with increasing demand (negative mode) or decreasing demand (positive mode). On the other hand, the extent of selection is dependent on the value of C and increases, reaches a maximum, and then declines as demand increases. The combination of these results suggests that the optimum extent and rate of selection occurs at around D=0.001 for the negative mode and 1-D=0.01 for the positive mode. In the case of the positive mode this represents a choice of 1-D that yields a rate of selection that is nearly equivalent to the optimum for the negative mode.

The allowed regions for selection permit one for the first time to specify precisely what is meant by high and low demand. With the nominal values for the parameters of the lactose and maltose operons in *E. coli*, selection of the negative mode of control requires a demand between 0.000005 and 0.1, whereas selection of the positive mode requires a demand between 0.2 and 0.999985. Furthermore, these regions exhibit the predicted asymmetry with the positive mode having the larger region within which selection is realizable.

The quantitative theory reveals a number of new relationships involving cycle time that can be tested against experimental data in the case of the lactose and maltose operons of *E. coli*. The first such relationship provides an estimate for the minimum value of the cycle time C_{min}. We obtained values of 26 hours and 10 hours, which is on the same order of magnitude as the 40 hours required on average for transit through the entire intestinal tract. Under these

circumstances, *E. coli* is simply passing through the intestinal tract without colonizing the colon. Clearly, the cycle time can be no shorter than this period.

The second relationship provides an estimate for the maximum value of the cycle time C_{max}. We have estimated this value to be approximately 580,000 hours (~66 years) in the case of the lactose operon and 502,000 hours (~ 57 years) in the case of the maltose operon. These values for C_{max} are on the same order of magnitude as the 120-year maximum for the life span of humans. Clearly, the cycle time for *E. coli* can be no longer than the life-time of the host because the bacteria will die with the host if they do not colonize a new host.

The final relationship provides an estimate for the optimal value of the cycle time C_{op}. The optimal extent and rate of selection determined for the lactose operon suggest a demand in the neighborhood of $D_{op} = 0.001$. This value of D, taken together with the measured three-hour exposure time to lactose in humans ($D=3/C$), predicts an optimal cycle time of $C_{op} = 3000$ hours (~ 4 months). The corresponding estimate based on the maltose operon is $C_{op} = 800$ hours (~ 33 days). These predicted values for the cycle time of *E. coli* are comparable with the cycle times (recolonization rates) of months to years that have been observed in humans for resident strains of *E. coli*.

In summary, the quantitative development of demand theory presented in the first paper and applied in the second provides the first estimates for the high and low values of demand that are required for selection of the positive and negative modes of gene control. The specific application to the maltose and lactose operons of *E. coli* suggests that the positive and negative modes of control for these genes are subject to selection throughout the full range of cycle times that are possible for this microbe. Moreover, the cycle times predicted on the basis of the optimal extent and rate of selection are in agreement with the typical cycle times that have been observed experimentally.

The quantitative version of demand theory integrates information at the level of DNA (mutation rate, effective target sizes for mutation of regulatory proteins, promoter sites, and modulator sites), physiology (selection coefficients for superfluous expression of an unneeded function and for lack of expression of an essential function), and ecology (environmental context and life cycle) and makes rather surprising predictions connected to the intestinal physiology and life span of the host and to the rate for recolonizing the host. Two additional approaches have yielded results that provide new insight into the normal physiological operation of this gene circuit in the organism's natural environment.

First, we have extended our initial analysis to include the AND logic by which expression of the lactose operon is determined by the absence of the lactose-specific repressor and the presence of the global CAP activator [4]. When the logic of combined control by CAP-cAMP activator and Lac repressor was analyzed, we found an optimum set of values not only for the exposure to lactose, but also for the exposure to glucose and for the relative phasing between the periods of exposure.

Second, we also have analyzed the alternative types of switching behavior that can be exhibited by the lactose operon in *Escherichia coli* [5]. The analysis showed that slight changes in the connectivity of the circuit (made possible by the used of metabolic analogues) can alter the switching behavior from a continuous graded response to a discontinuous all-or-none expression of the *lac* operon.

Distinguish Design Features That Occur as the Result of Natural Selection from Those That Occur With High Probability by Chance

We have been comparing alternative designs for achieving the same function in an effort to identify those designs that would arise frequently by chance and that would be robust in the face of mutational entropy. Any method to identify the critical differences between alternative designs for gene circuitry must be able to distinguish those design features that are the result of natural selection from those that occur with high probability by chance. The principal method we have used for most of our studies is the method of mathematically controlled comparisons.

The method involves the following steps. (1) The alternatives being compared are restricted to having differences in a single specific process that remains embedded within its natural milieu. (2) The values of the parameters that characterize the unaltered processes of one system are assumed to be strictly identical with those of the corresponding parameters of the alternative system. This equivalence of parameter values within the systems is called *internal equivalence.* It provides a means of nullifying or diminishing the influence of the background, which in complex systems is largely unknown. (3) Parameters associated with the changed process are initially free to assume any value. This allows the creation of new degrees of freedom. (4) The extra degrees of freedom are then systematically reduced by imposing constraints on the external behavior of the systems. In this way, the two systems are made as nearly equivalent as possible in their interactions with the outside environment. This is called *external equivalence.* (5) The constraints imposed by external equivalence fix the values of the altered parameters in such a way that arbitrary differences in systemic behavior are eliminated. Functional differences that remain between alternative systems with maximum internal and external equivalence constitute irreducible differences. (6) When all degrees of freedom have been eliminated, and the alternatives are as identical as they can be, then comparisons are made by rigorous mathematical and computer analyses of the alternatives.

In some cases, the results obtained using this technique are general and independent of parameter values and the answers are clear-cut. In others, the result might be general, but the demonstration is difficult and numerical results with specific parameter values can help to clarify the situation. In either situation, numerical results with specific parameter values also can provide an answer to the question of how much better one of the alternatives might be. In contrast, a more ambiguous result is obtained when either of the alternatives can have the larger value for a given systemic property, depending on the specific values of the parameters. In any case, introduction of specific values for the parameters reduces the generality of the results. A numerical approach to this problem has been developed that combines the method of mathematically controlled comparison with statistical techniques to yield numerical results that are general in a statistical sense. These developments have been reported in a series of five papers [6-10].

The first task in developing a statistical version of mathematically controlled comparisons was to expand the usual methodology of making statistical comparisons. When dealing with questions that concern a general class of models for biological networks, large numbers of distinct models within the class can be grouped into an ensemble that gives a statistical view of the properties for the general class. However, comparing properties of different ensembles through the use of the usual point measures (e.g. medians, standard deviations, correlation coefficients) can mask inhomogeneities in the correlations between properties. We were therefore motivated to develop strategies that allow these inhomogeneities to be more easily detected [6]. First, we take advantage of the regular systematic structure of the power-law formalism to construct ensembles of parameter sets for both the reference model and the alternative model in question. Second, these realizations of the two alternative model designs are analyzed to characterize their systemic behaviors. Third, these are then compared by means of a novel "Density of Ratios Plot". Techniques involving moving quantiles are introduced to generate secondary plots in which correlations and inhomogeneities in correlations are more easily detected. Finally, we provided several examples to illustrate the advantages of these techniques.

The first uses of the graphical and statistical methods presented in the previous paper were to examine how the different systemic properties of a biochemical network are correlated with one another and how the specification of particular systemic properties biases the distribution of the underlying parameter values [7]. To keep the application as transparent as possible we examined a simple unbranched biosynthetic pathways subject to control by feedback inhibition. After constructing a large ensemble of randomly generated sets of parameter values, the structural and behavioral properties of the model with these parameter sets were examined statistically and classified. The results of our analysis demonstrated that certain properties of these systems are strongly correlated, thereby revealing aspects of organization that are highly probable independent of selection. This is an aspect of the usual approach whereby parameter values are examined to determine their influence on systemic behavior. We have also taken the opposite view to learn how selection for particular systemic behaviors influences the frequency distribution of parameter values. Information on the distribution of parameter values is of interest in the design of experiments to measure the parameters in actual systems. By knowing the most probable values of a parameter, one can design experiments to target that range. The results would allow us to make predictions about the range of values most likely to generate systems with the known behavior.

Having developed the statistical methods and examined the correlations among the properties of a system, we used these methods for the specific purpose of extending the method of Mathematically Controlled Comparison [8]. This method has been used for some time to determine which of two alternative regulatory designs is better according to specific quantitative criteria for functional effectiveness. We were motivated to develop and apply statistical methods that would permit the use of numerical values for the parameters and yet retain some of the generality that makes Mathematically Controlled Comparison so attractive. We illustrated this new numerical method in a step-by-step application using a very simple didactic example. We also validated the results by comparison with the corresponding results obtained using the previously developed analytical method. The numerical method confirmed the qualitative differences between the systemic behavior of alternative designs obtained from the analytical

method. In addition, the numerical method allowed for quantification of the differences and it provided results that are general in a statistical sense.

The statistical strategy for making mathematically controlled comparisons, after having been developed and validated, was used to address two long-standing biological question: (1) Why is the pattern of overall feedback inhibition a nearly universal design feature of unbranched biosynthetic pathways? (2) Why do most unbranched biosynthetic pathways have irreversible reactions near their beginning, many times at the first step?

We approached the first of these questions by examining pathways with an arbitrary pattern of feedback interactions and an otherwise equivalent pathway without the overall feedback interaction [9]. Our statistical method allowed the rigorous determination of the changes in systemic properties that can be exclusively attributed to overall feedback inhibition. Analytical results show that the unbranched pathway can achieve the same steady-state flux, concentrations and logarithmic gains with respect to changes in substrate, with or without overall feedback inhibition. The analytical approach also shows that overall feedback inhibition amplifies the regulation of flux by the demand for end product while attenuating the sensitivity of the concentrations to the same demand. This approach does not provide a clear answer regarding the effect of overall feedback inhibition on the robustness, stability and transient time of the pathway. However, the generalized numerical method we used does clarify the answers to these questions. On average, an unbranched pathway with overall feedback inhibition is less sensitive to perturbations in the values of the parameters that define the system. On average, overall feedback inhibition decreases the stability margins by a minimal amount (typically less than 5%). Finally, and again on average, stable systems with overall feedback inhibition respond faster to fluctuations in the metabolite concentrations. Taken together these results show that overall feedback inhibition confers several functional advantages upon unbranched pathways. These advantages provide a rationale for the prevalence of this mechanism in unbranched metabolic pathways *in vivo*.

We approached the second of these questions by systematically varying the position of the irreversible reaction in model pathways and comparing the systemic behavior according to several criteria for functional effectiveness using the method of mathematically controlled comparisons [10]. This technique minimizes extraneous differences in systemic behavior and identifies those that are fundamental. Our results show that a pathway with an irreversible reaction located at the first step, and all other reactions reversible, is on average better than an otherwise equivalent pathway with all reactions reversible, which in turn is on average better than an otherwise equivalent pathway with an irreversible reaction located at any step other than the first. Pathways with an irreversible first reaction and low concentrations of intermediates (one of the primary criteria for functional effectiveness) exhibit the following profile when compared to fully reversible pathways: changes in the concentration of intermediates in response to changes in the level of initial substrate are equally low, the robustness of the intermediate concentrations and of the flux is similar, the margins of stability are similar, flux is more responsive to changes in demand for end product, intermediate concentrations are less responsive to changes in demand for end product, and transient times are shorter. These results provide a functional rationale for the positioning of irreversible reactions at the beginning of unbranched biosynthetic pathways.

Significance / Impact / Navy Relevance:

While it is obvious that gene circuits are capable of performing various computational tasks, it is not clear why particular circuitry has evolved to perform these tasks. If we can develop a deeper understanding of the basic design principles, significant biological consequences would follow. Moreover, we might be able to supply an appropriate context for the directed evolution of robust circuits with desirable computational properties, which could also have important technological ramifications.

Several elements of design, each exhibiting a variety of realizations, have been identified among elementary gene circuits in prokaryotic organisms. Design principles that appear to govern the realization for some of these elements have been identified by the use of well-controlled mathematical comparisons. Work on this grant has contributed to the identification of four such principles. These make specific predictions regarding (1) two alternative modes of gene control, (2) three patterns of coupling gene expression in elementary circuits, (3) two types of switches in inducible gene circuits, and (4) the realizability of alternative gene circuits and their response to phased environmental cues. In each case, the predictions are supported by experimental evidence. These results are important for understanding the normal function of gene circuits; they also are potentially important for developing judicious methods to redirect normal expression for biotechnological purposes or to correct pathological expression for therapeutic purposes.

In this third phase of our work we have begun to address the task of using directed evolution to design circuits that perform specific functions. The statistical methods we developed have helped us identify designs that meet specific performance criteria by starting with random assignment of parameter values and applying selection. The resulting designs that occur with high probability are likely to be robust. If this approach proves successful for a wide variety of circuits performing diverse functions, then it would provide the basis for experimental approaches aimed at a biological realization of these circuits.

A review covering all of the work undertaken during the period of this grant has recently been published [11]. This work has motivated us to continue our investigation of randomly-generated gene circuits and to focus our attention on the role of alternative forms of connectivity. The first paper representing this new work in progress is currently under revision [12].

References:

[1] Savageau, M.A., Rules for the evolution of gene circuitry, *in* "Pacific Symposium on Biocomputing '98," R.B. Altman, A.K. Dunker, L. Hunter and T.E. Klein, Eds., World Scientific, Singapore, pp. 54-65 (1998).
[2] Savageau, M.A., Demand theory of gene regulation: Quantitative development of the theory, Genetics **149**, 1665-1676 (1998).
[3] Savageau, M.A., Demand theory of gene regulation: Quantitative application to the lactose and maltose operons of *Escherichia coli*, Genetics **149**, 1677-1691 (1998).

[4] Savageau, M.A., Evolution and regulation of gene circuitry: Functional genomics of the lactose pathway in *E. coli*, *in* "Proceedings of the Workshop on Computation of Biochemical Pathways and Genetic Networks," E. Bornberg-Bauer, A. De Beuckelaer, U. Kummer, and U. Rost, Eds., Logos Verlag, Berlin, Germany, pp. 57-66 (1999).
[5] Savageau, M.A., Design of gene circuitry by natural selection: Analysis of the lactose catabolic system in *E. coli*, Biochem. Soc. Trans. **27**, 264-270 (1999).
[6] Alves, R., and M.A. Savageau, Comparing systemic properties of ensembles of biological networks by graphical and statistical methods, Bioinformatics **16**, 527-533 (2000).
[7] Alves, R., and M.A. Savageau, Systemic properties of ensembles of metabolic networks: application of graphical and statistical methods to simple unbranched pathways, Bioinformatics **16**, 534-547 (2000).
[8] Alves, R., and M.A. Savageau, Extending the method of mathematically controlled comparison to include numerical comparisons, Bioinformatics **16**, 786-798 (2000).
[9] Alves, R., and M.A. Savageau, Effect of overall feedback inhibition in unbranched biosynthetic pathways, Biophysical J. **79**, 2290-2304 (2000).
[10] Alves, R. and M.A. Savageau, Irreversibility in unbranched pathways: preferred positions based on regulatory considerations, Biophysical J. **80**, 1174-1185 (2001).
[11] Savageau, M.A., Design principles for elementary gene circuits: Elements, methods, and examples, *Chaos* **11**, 142-159 (2001).
[12] Oosawa, C. and M.A. Savageau, Effects of alternative connectivity on behavior of randomly constructed genetic networks, Physica D **000**, 0000-0000 (2001).

PACIFIC SYMPOSIUM ON BIOCOMPUTING '98

Maui, Hawaii, USA
4–9 January 1998

Editors

Russ B. Altman
Stanford University

A. Keith Dunker
Washington State University

Lawrence Hunter
National Institutes of Health

Teri E. Klein
University of California, San Francisco

World Scientific
Singapore • New Jersey • London • Hong Kong

RULES FOR THE EVOLUTION OF GENE CIRCUITRY

M.A. SAVAGEAU
Department of Microbiology & Immunology, The University of Michigan Medical School, Ann Arbor, Michigan 48109-0629 USA

Cells possess the genes required for growth and function in a variety of contexts. In any given context there is a corresponding pattern of gene expression in which some genes are OFF and others ON. The ability of cells to switch genes ON and OFF in a coordinate fashion to produce the required patterns of expression is the fundamental basis for complex processes like normal development and pathogenesis. The molecular study of gene regulation has revealed a plethora of mechanisms and circuitry that have evolved to perform what appears to be the same switching function. To some this implies the absence of rules. However, simple rules capable of relating molecular design to the natural environment have begun to emerge through the analysis of elementary gene circuits. Two of these rules are reviewed in this paper. These simple rules have the ability to unify understanding across several different levels of biological organization -- molecular, physiological, developmental, ecological.

1. Introduction

Regulation of gene expression and its systemic manifestations are subjects of intense study. As a result of this effort we shall soon have identified all of the genes and proteins for a number of simpler organisms. Despite this enormous progress we are still at a loss to understand the integrated behavior of the organism. Our understanding is still fragmentary and descriptive. We are unable to predict changes in the organism's behavior when it is placed in a novel environment or when a change is made in one of its genes. Little is known about the forces that lead to the selection or maintenance of a specific mechanism for the regulation of a given set of genes in a particular organism. Is this process random, or is it governed by rules? The answer to this question is important. It will help us to understand the evolution of gene regulation; it also will help us to develop judicious methods of redirecting normal expression for biotechnological purposes or of correcting pathological expression for therapeutic purposes.

Our goal is to understand the integrated structure and function of organizationally complex systems in relation to their underlying molecular determinants. Moreover, we are particularly interested in identifying the rule-like properties of these systems that would allow for some algorithmic compression in their representation, and not simply a compilation of all the molecular details.

In pursuit of this goal we have developed a canonical nonlinear formalism that has desirable properties for the representation and analysis of organizationally complex systems (1). This formalism has been used to characterize alternative

modes of gene control and various forms of coupling among elementary gene circuits. The results allow us to identify a set of rules, or design principles, that govern the natural selection of gene circuits. Here we shall review the relevant biological background and then present results from our analysis of gene circuitry.

2. Biological Background

The common metaphor of the genome as a blueprint for construction of the organism masks the difficult task of relating structure and function of the intact organism to its underlying genetic determinants (2). The behavior of an intact biological system can seldom be related directly to its underlying molecular determinants. There are several different levels of hierarchical organization that are relevant. For our present purposes it will be sufficient to consider four different levels -- genome sequence, transcriptional unit, elementary gene circuit, environmental context.

2.1. The DNA sequence constitutes the genome

The recent sequencing of the complete genome for a number of simpler organisms, and the projected completion of the sequence for the human genome by the year 2005, illustrate the power of modern molecular biology to resolve complex systems into their simplest elements. The four bases -- A, T, G, and C -- are strung together in sequences that are mind-numbing in their simplicity; yet, these sequences provide the potential for incredible complexity. Whether it be the versatile metabolism of free-living microbes that can adapt to nearly any environment, or the sophisticated structures of multicellular organisms that can be seen in near endless variety, the physical basis for this complexity is the context-dependent expression of the organism's genome.

2.2. Information is encoded in transcriptional units

The mapping from DNA level to organismal level requires a deeper understanding of how information is encoded in the genome. DNA sequences are organized into functional units that consist of structural genes flanked by a start sequence at which transcription begins and a termination sequence at which it ends. In addition, there are a number of regulatory sites capable of binding specific transcription factors that interact with the transcription machinery to modulate the rate of transcription initiation or termination (Fig. 1).

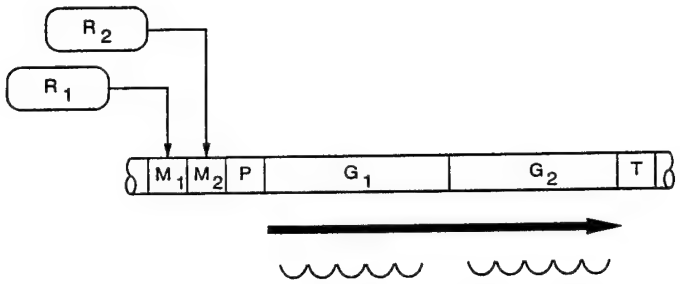


Figure 1. Unit of transcription. Structural genes (G_1 and G_2) are bounded by a promoter sequence (P) and a terminator sequence (T), and preceded by upstream modulator sites (M_1 and M_2) that bind regulators (R_1 and R_2) capable of altering transcription initiation. The solid arrow represents the mRNA transcript and the scalloped lines indicate the protein products encoded by genes G_1 and G_2.

2.3. Expression is organized into elementary gene circuits

Transcription of DNA is but one step in a cascade of information flow that constitutes the expression of a gene (Fig. 2). Each stage of such a cascade is a potential site at which expression can be regulated in a context-dependent fashion. The context is provided by the life cycle of the organism, and the interlocking mechanisms of gene regulation interpret that context.

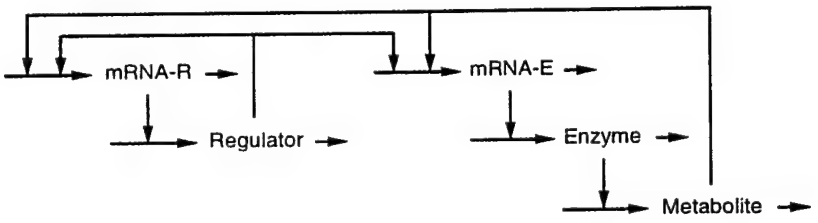


Figure 2. Cascade of information flow from DNA to RNA to protein to metabolite. The processes of synthesis and degradation are represented by horizontal arrows, whereas the catalytic and regulatory influences are represented by vertical arrows. An effector circuit is shown on right and a regulator circuit is shown on the left.

2.4. Physiology and ecology are reflected in the organism's life cycle

The life cycle of some organisms is largely programmed development from egg to embryo to mature adult and back to the egg (3). In other organisms it is dominated by random events involving a pathogen's ability to encounter one host, to exploit or

colonize that host for a period of time, to escape into a secondary environment, and to survive there until an encounter with a subsequent host (4). In each case, specific genes function in some phases of the organism's life cycle but not in others. Differential patterns of expression are exhibited as the context changes from one phase to the next and one set of genes is switched OFF while another set is switched ON in a combinatorial fashion.

Gene regulation -- the ability to switch gene expression ON and OFF in appropriate temporal and spatial patterns -- is central to modern biology. The inability to express a gene when it should be ON, or the inappropriate expression of a gene when it should be OFF, is usually dysfunctional and often lethal. The determination of what constitutes appropriate expression requires knowledge of the molecular mechanism, the physiological function it realizes, and the environmental demand for that function.

Organisms regulate expression of their genome by means of a diverse repertoire of molecular mechanisms. Most of the well-characterized examples have come from the study of prokaryotes. Although the situation is typically more complex in eukaryotes and there are undoubtedly some aspects of regulation unique to higher organisms, the general themes are much the same in both and most mechanisms that were originally thought to be unique to eukaryotes have subsequently been observed within the prokaryotic realm. For our analysis, we have abstracted the generic features of gene regulation that are thought to be common to both, but for testing our predictions we have turned to the more numerous and well-characterized prokaryotes systems. The extent to which the results might differ for eukaryotes remains to be determined.

3. Rules for the Molecular Mode of Gene Control

One of the first variations in design to be well documented is that involving positive vs. negative modes of gene control (Fig. 3). For example, the lactose (*lac*) catabolic system in *Escherichia coli* is governed by a classical repressor protein (5), the negative mode of control. Induction of gene expression in this system is achieved by the addition of an inducer that removes the repressor protein to allow transcription. The maltose (mal) system in *E. coli*, by contrast, is governed by an activator protein (6), the positive mode of control. Induction in this case is achieved by the addition of an inducer that converts the activator protein into its functional form that facilitates transcription. What is the significance of this variation in design?

This difference in design was originally believed to have no functional significance. Subsequent analysis showed that mode of control is related to the of control showed that in most respects their behavior can be identical. However,

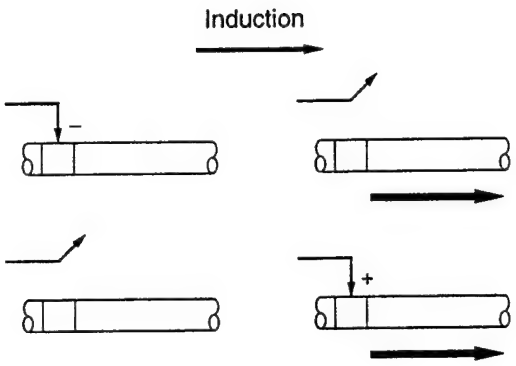


Figure 3. Alternative molecular modes of controlling gene expression.

demand for expression of the regulated gene in the organism's natural environment (7). The analysis of mathematical models with either the positive or negative mode they behave in diametrically opposed ways to mutations in the components of the regulatory mechanism itself. Mutants altered in the positive mechanism are unable to express the corresponding gene product despite the presence of inducer, whereas mutants altered in the negative mechanism express the corresponding gene product even in the absence of inducer. The relative growth of mutant and wild-type organisms was examined in high- and low-demand environments. The high-demand environment, in which high-level expression is frequently required for the organism's survival, leads to selection of the positive mode of gene control; the low-demand environment leads to selection of the negative mode. Thus, molecular mode of control is correlated with level of demand for expression of the regulated gene product in the organism's natural environment (Table 1). These qualitative predictions are well supported by experimental evidence (8).

Table 1. Predicted correlation between demand for expression and mode of control

Demand for expression	Mode of regulation	
	Positive	Negative
High	Regulation selected	Regulation lost
Low	Regulation lost	Regulation selected

In recent analysis we have examined the quantitative implications of this demand theory (in preparation). First, we define two key parameters: the cycle time C, which is the average time for a gene to cycle through the OFF state, the ON state, and back to the OFF state; and demand D, which is the fraction of the cycle time that the gene is ON. Second, a quantitative analysis involving mutation rates and growth rates reveals non-overlapping regions in the C vs. D space for which selection of wild-type regulatory mechanisms with the negative or the positive mode is realizable (Fig. 4).

The quantitative theory specifies more precisely what we mean by high and low demand. As can be seen in Figure 4, with the nominal values for the parameters of the lactose and maltose operons in *E. coli*, selection of the negative mode of control requires a demand less than 0.04, whereas selection of the positive mode requires a demand greater than 0.32.

Although these limits on demand are influenced by a number of parameters, by far the most influential parameter is the reduction in growth rate when there is excess expression of a gene whose function is not required. The nominal value for

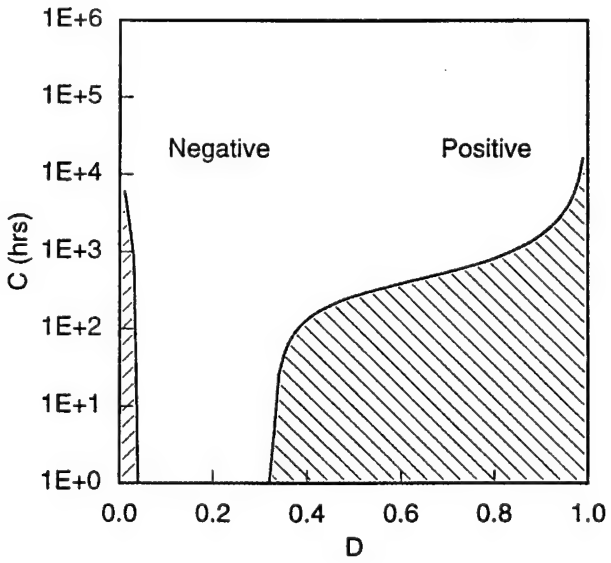


Figure 4. Thresholds for discriminate selection of wild-type regulatory mechanisms with negative or positive modes. There is maximum value of demand for selection of the negative mode and a minimum value of demand for selection of the positive mode. The values of cycle time C and demand D are based on parameter values for the *lac* and *mal* systems in *E. coli*.

this parameter was set at 5%, based on data for the lactose operon that suggest this value as a maximum for the reduction in growth rate of operator-constitutive mutants in a low-demand environment. In the case of the positive mode, the same value was used to characterize the reduction in growth rate of an up-promoter mutant in a low-demand environment. A 10% variation in this parameter yields a two-fold change in the limits of D for both the negative and positive mode. The remaining parameters have much less influence on the limits of D; approximately half exhibit a nearly linear influence, whereas the other half have a negligible influence.

4. Rules for the Coupling of Elementary Gene Circuits

A second variation in design is that involving the coupling of elementary gene circuits for regulator and effector genes. Early experimental studies (9) suggested that expression of regulator genes is invariant in some cases (classical regulation), such as in the *lac* system in *E. coli*, and coordinate with the regulated effector genes in other cases (autogenous regulation), such as in the histidine utilization (*hut*) system in *Salmonella typhimurium*. Our earlier work focused on the functional implications of these alternatives, which we now refer to as the completely uncoupled and perfectly coupled patterns of regulator and effector gene expression (10). However, inducible systems with other patterns of gene expression were subsequently reported, and these have become the stimulus to extend our earlier work.

Logically, there are three qualitatively distinct patterns of regulator and effector gene expression that can be exhibited by an inducible system (Fig. 5). These are the directly coupled, uncoupled, and inversely coupled patterns in which regulator gene expression increases, remains the same, and decreases with an increase in effector gene expression. Well-studied examples of direct coupling, uncoupling, and inverse coupling are provided by the D-serine deaminase (11), arabinose (6), and methionine (12) systems in *E. coli*.

The functional implications of direct coupling, uncoupling, and inverse coupling have been determined from an analysis of a generalized model capable of representing these different forms of coupling (Fig. 6). The fundamental equations that characterize this model are mass-balance equations that take the general form

$$dX_i/dt = V_{+i}(X_1, \ldots, X_8) - V_{-i}(X_1, \ldots, X_8) \qquad i = 1, \ldots, 5 \qquad (1)$$

The rate laws V_{+i} and V_{-i} describe mass fluxes due to synthetic and degradative processes. These rate laws can be represented as products of power-law functions according to the results of theoretical analyses (1) and empirical case studies (13). Thus, we can rewrite Eq. 1 to obtain the following system of equations:

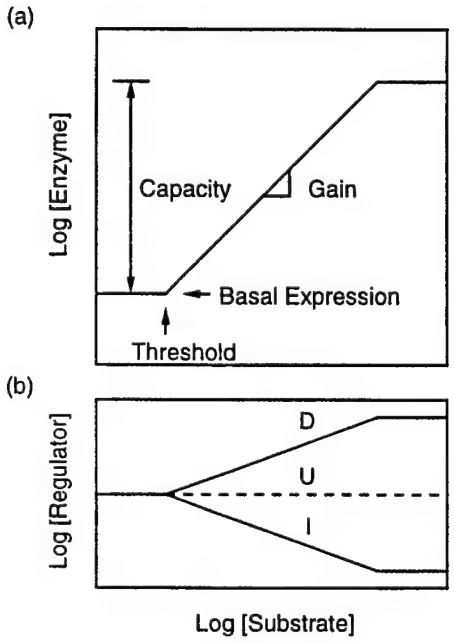


Figure 5. Expression characteristics for (a) effector and (b) regulator gene expression. Three distinct patterns of coupling are illustrated. Effector gene expression increases while regulator gene expression (D) increases (directly coupled), (U) remains unchanged (uncoupled), or (I) decreases (inversely coupled).

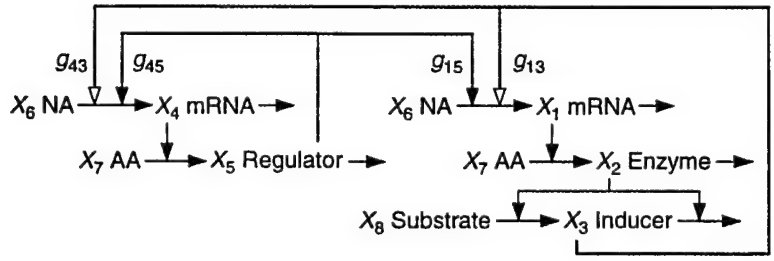


Figure 6. Coupled circuits for the expression of regulator and effector genes. Mass fluxes that characterize the state of the system are represented by horizontal arrows, whereas catalytic and regulatory influences are represented by vertical arrows. The influences of the regulator (closed arrowheads) are described by the kinetic orders $g15$ and $g45$; the influences of the inducer (open arrowheads) are described by the kinetic orders $g13$ and $g43$ (see Eqs. 2-6).

$$dX_1/dt = V_{+1} - V_{-1} = \alpha_1 X_6^{g16} X_3^{g13} X_5^{g15} - \beta_1 X_1^{h11} \quad (2)$$

$$dX_2/dt = V_{+2} - V_{-2} = \alpha_2 X_7^{g27} X_1^{g21} - \beta_2 X_2^{h22} \quad (3)$$

$$dX_3/dt = V_{+3} - V_{-3} = \alpha_3 X_8^{g38} X_2^{g32} - \beta_3 X_2^{h32} X_3^{h33} \quad (4)$$

$$dX_4/dt = V_{+4} - V_{-4} = \alpha_4 X_6^{g46} X_3^{g43} X_5^{g45} - \beta_4 X_4^{h44} \quad (5)$$

$$dX_5/dt = V_{+5} - V_{-5} = \alpha_5 X_7^{g57} X_4^{g54} - \beta_5 X_5^{h55} \quad (6)$$

These equations are used to analyze systems with the positive or negative mode of control for each circuit. The effects of physicochemical limitations, which arise from the subunit structure of regulator proteins and place bounds on kinetic orders in this model (10), are also considered. The functional effectiveness of these various circuits has been compared on the basis of several properties (decisiveness, efficiency, selectivity, robustness, stability, and responsiveness) that represent possible criteria for natural selection. Of these, responsiveness has proved the most sensitive to variations in circuit design (14).

The results allow us to predict a correlation between the form of coupling and the capacity for induction (ratio of maximal to minimal level of effector gene expression). Negatively controlled systems with low, intermediate, and high capacities for gene expression are predicted to have direct coupling, uncoupling, and inverse coupling, respectively. Positively controlled systems, in contrast, are predicted to have inverse coupling, uncoupling, and direct coupling (Table 2).

These predictions are compared with data available in the literature for systems in which the pattern of regulator and effector gene expression is known (Fig. 7). They are found to be in reasonable agreement, given measurement error.

Table 2. Predicted correlation between circuitry and capacity for regulation

Demand	Mode	Capacity	Circuit
High	Positive	Low	Inversely coupled
High	Positive	Intermediate	Uncoupled
High	Positive	High	Direct coupled
Low	Negative	Low	Directly coupled
Low	Negative	Intermediate	Uncoupled
Low	Negative	High	Inversely coupled

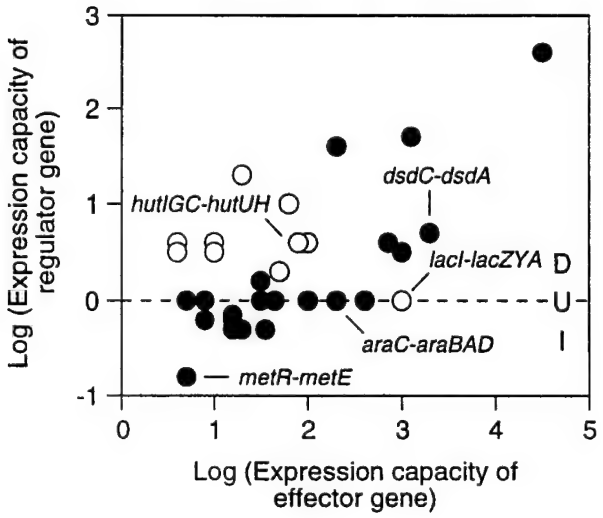


Figure 7. Patterns of regulator and effector gene expression in inducible systems of bacteria. Expression of each gene is measured in the presence of excess inducer and normalized with respect to its basal level in the absence of inducer. For the effector gene this is equivalent to its capacity; for the directly coupled regulator gene this also is equivalent to its capacity, but for the inversely coupled regulator gene this is equivalent to the inverse of its capacity. Estimates of capacity are based on published reports. Directly coupled (D), uncoupled (U), and inversely coupled (I) systems are represented above, on, and below the dashed line, respectively. Negatively regulated systems are shown as open circles; positively regulated systems are shown as closed circles.

5. Discussion

The genome of an organism evolves to realize a developmental program with specific gene circuitry that can be viewed as computing the solution to the environmental problem faced by the organism. This is a suggestive metaphor, but at present we have little understanding of the circuits and the computations they might perform. The large number of genes encoded in the DNA of even the simplest of organisms suggests that this circuitry might be very complex and exhibit a high degree of connectivity. If this were the case, then the task of elucidating the circuitry would be daunting.

However, a number of different lines of evidence suggest that although there may be a large number of gene circuits, they may have a minimal degree of connectivity. First, molecular analysis of gene regulation in bacteria has shown that most gene circuits are governed by a small number of regulators, usually one to

three. In eukaryotes the numbers are larger in some cases, but seldom more than a dozen regulators influence a given gene circuit. Second, the enumeration of regulators and their targets, based on sequence homologies, has shown the same results for bacteria; namely, one or two regulators affecting a given circuit (15,16). Third, computer simulations of large, randomly-connected circuits have been used to explore the question of connectivity. The most biologically-suggestive behaviors were found when each circuit was subject to two or three regulatory interactions, and less relevant behaviors were found with higher or lower degrees of connectivity (17).

Low degrees of connectivity suggest that a 'bottom-up' strategy of characterizing genome circuitry in terms of rules for elemental gene circuits is likely to prove fruitful. Indeed, this seems to be the case with our initial experience attempting to generalize on the basis of the few rules that we have uncovered to date. To give one example, consider the carbon regulation system in *E. coli*.

Carbon regulation in *E. coli* is manifested in large part through the action of the cyclic AMP receptor protein (CRP)-cyclic AMP (cAMP) system (18), which was among the first global regulators to be characterized. This system coordinates the utilization of diverse sources of carbon whose levels vary in both time and space. An application of demand theory indicates that all of the regulators in this system fit a self-consistent pattern. Because the CRP-cAMP regulator is an activator of transcription for the inducible catabolic systems, one can predict that at least some of these systems are in high demand in the organism's natural environment. Indeed, a number of the inducible systems for non-PTS substrates are controlled by specific activators (8). Conversely, one can predict that the PTS substrates, which repress the levels of CRP-cAMP, are seldom present in high concentrations in the natural environment. Indeed, all of the inducible systems for PTS substrates that have been examined involve control by a specific repressor (8), which again is what one would predict according to demand theory. Thus, at least the modality of all the regulators in this system seem to be self-consistent.

In conclusion, regulation of gene expression is clearly one of the most fundamental processes in the living world. Knowledge of gene regulation is a prerequisite for understanding function, adaptation and evolution, and such understanding will in turn be essential for the design and implementation of novel metabolic pathways by means of genetic engineering. The results of our studies suggest that although there is an enormous diversity of mechanisms, there also are well-established patterns that can be understood in terms of simple rules.

Acknowledgments

This work was supported in part by grants from the U.S. National Institutes of Health and Office of Naval Research.

References

1. M.A. Savageau, in *World Congress of Nonlinear Analysts 92*, Vol. 4, Ed. V. Lakshmikantham (Walter de Gruyter Publishers, Berlin, 1996), pp. 3323-3334
2. F.C. Neidhardt and M.A. Savageau, in *Escherichia coli and Salmonella; cellular and molecular biology*, Vol. 1, 2nd ed., Eds. F. C. Neidhardt, et al. (ASM Press, Washington, D.C., 1996), pp. 1310-1324
3. J.M.W. Slack, *From Egg to Embryo*, 2nd ed. (Cambridge University Press, Cambridge, 1992)
4. A.A. Salyers, *Bacterial Pathogenesis: A Molecular Approach* (ASM Press, Washington, D.C., 1994)
5. H. Choy and S. Adhya, in *Escherichia coli and Salmonella; cellular and molecular biology*, Vol. 1, 2nd ed., Eds. F. C. Neidhardt, et al. (ASM Press, Washington, D.C., 1996), pp. 1287-1299
6. R. Schleif, in *Escherichia coli and Salmonella; cellular and molecular biology*, Vol. 1, 2nd ed., Eds. F. C. Neidhardt, et al. (ASM Press, Washington, D.C., 1996), pp. 1300-1309
7. M.A. Savageau, *Proc. Nat. Acad. Sci. USA* **74**, 5647-5651 (1977)
8. M.A. Savageau, in *Theoretical Biology -- Epigenetic and Evolutionary Order*, Eds. B.C. Goodwin and P.T. Saunders (Edinburgh University Press, Edinburgh, 1989), pp. 42-66
9. R.F. Goldberger, *Science* **183**, 810-816 (1974)
10. W.S. Hlavacek and M.A. Savageau, *J. Mol. Biol.* **248**, 739-755 (1995)
11. E. McFall, in *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*, Eds. F.C. Neidhardt et al. (American Society for Microbiology, Washington, D.C., 1987), pp. 1520-1526
12. I.G. Old, S.E. Phillips, P.G. Stockley and I. Saint Girons, *Prog. Biophys. Mol. Biol.* **56**, 145-185 (1991)
13. E.O. Voit, *Canonical Nonlinear Modeling: S-System Approach to Understanding Complexity* (Van Nostrand Reinhold, New York, 1991)
14. W.S. Hlavacek and M.A. Savageau, *J. Mol. Biol.* **255**, 121-139 (1996)
15. J. Collado-Vides, B. Magasanik, and J.D. Gralla, *Microbiol. Rev.* **55**, 371-394 (1991)
16. J. Otsuka, H. Watanabe, and K.T. Mori, *J. Theoret. Biol.* **178**, 183-204 (1996)
17. S.A. Kauffman, *J. Theoret. Biol.* **22**, 437-467 (1969)
18. A. Kolb, S. Busby, H. Buc, S. Garges, and S. Adhya, *Annu. Rev. Biochem.* **62**, 749-795 (1993)

Copyright © 1998 by the Genetics Society of America

Demand Theory of Gene Regulation. I. Quantitative Development of the Theory

Michael A. Savageau

Department of Microbiology and Immunology, The University of Michigan, Ann Arbor, Michigan 48109-0620

Manuscript received December 15, 1997
Accepted for publication May 6, 1998

ABSTRACT

The study of gene regulation has shown that a variety of molecular mechanisms are capable of performing this essential function. The physiological implications of these various designs and the conditions that might favor their natural selection are far from clear in most instances. Perhaps the most fundamental alternative is that involving negative or positive modes of control. Induction of gene expression can be accomplished either by removing a restraining element, which permits expression from a high-level promoter, or by providing a stimulatory element, which facilitates expression from a low-level promoter. This particular design feature is one of the few that is well understood. According to the demand theory of gene regulation, the negative mode will be selected for the control of a gene whose function is in low demand in the organism's natural environment, whereas the positive mode will be selected for the control of a gene whose function is in high demand. These qualitative predictions are well supported by experimental evidence. Here we develop the quantitative implications of this demand theory. We define two key parameters: the cycle time C, which is the average time for a gene to complete an ON/OFF cycle, and demand D, which is the fraction of the cycle time that the gene is ON. Mathematical analysis involving mutation rates and growth rates in different environments yields equations that characterize the extent and rate of selection. Further analysis of these equations reveals two thresholds in the C *vs.* D plot that create a well-defined region within which selection of wild-type regulatory mechanisms is realizable. The theory also predicts minimum and maximum values for the demand D, a maximum value for the cycle time C, as well as an inherent asymmetry between the regions for selection of the positive and negative modes of control.

DIFFERENTIAL regulation of gene expression is central to much of modern biology. Animal development can be thought of in terms of an early phase, which begins with an egg and ends with an embryo, and a late phase, which begins with an embryo and ends with the mature organism (SLACK 1992). Some genes function only in the early phase while others only in the late phase. The inability to express a gene when it should be ON or the excess expression of a gene when it should be OFF is usually dysfunctional and often lethal. For any given gene, expression can be considered a roughly periodic function, which in the simplest case is OFF for a period and ON for another period with the total duration being the lifetime of the organism. The differential regulation of many such genes in time and space determines the pattern of cell-specific expression that underlies development of the organism.

The life cycle of a bacterial association with a host organism also can be thought of in terms of an early phase, which begins with entry into a host organism and ends with successful colonization, and a late phase, which begins with colonization and ends, after a period of stable association, with the entry of another host (SALYERS 1994). Some bacterial genes function only in the early phase of initial colonization while others only in the late phase of stable association. Again, the inability to express a gene when it should be ON or the excess expression of a gene when it should be OFF is dysfunctional and in some cases lethal. Expression of any given gene is OFF for a period and ON for another period with the total duration in this case being the time for the bacteria to cycle from one host to another. Although the organisms in these two examples are quite different, in each case appropriate differential regulation of gene expression is clearly key to their survival.

A great deal is known about the molecular details of many gene systems, particularly in well-studied prokaryotic organisms. The wealth of studies in this area has revealed a variety of designs for the regulation of gene expression. However, we are just beginning to understand the functional implications of these various designs and to grasp the factors that have influenced their evolution.

One of the first variations in molecular design to be addressed was negative *vs.* positive modes for controlling gene expression. For example, the lactose (*lac*) operon in *Escherichia coli* is an inducible system with a negative mode of control by a repressor protein, the *lacI* gene product (MILLER and REZNIKOFF 1980). In an appropriate environment, induction occurs in response

Address for correspondence: 5641 Medical Science Bldg. II, Department of Microbiology and Immunology, The University of Michigan Medical School, Ann Arbor, MI 48109-0620. E-mail: savageau@umich.edu

to addition of the specific inducer, which results in removal of repressor and initiation of transcription. In contrast, the maltose (*mal*) operon is an inducible system with a positive mode of control by an activator protein, the *malT* gene product (SCHWARTZ 1987). Induction in this case involves the specific inducer binding to the activator protein, which is then able to interact with RNA polymerase and facilitate initiation of transcription. The same physiological function, induction, is being realized in each of these cases, but by alternative molecular mechanisms. Are these alternative designs historical accidents that are functionally equivalent, or have they been selected in nature because they exhibit functional differences?

An answer to this question was provided by demand theory (SAVAGEAU 1974, 1977, 1983a, 1989), which is based on selectionist arguments. In its simplest form, the theory can be understood in familiar qualitative terms and leads to the following predictions: a negative mode of control will be selected when there is a low demand for expression of the effector genes in the organism's natural environment; a positive mode will be selected when there is a high demand for their expression. These predictions, and a number of others that follow as natural extensions, have been tested in over 100 cases and there has been excellent agreement (SAVAGEAU 1979, 1983b, 1985).

Here I develop the quantitative implications of demand theory. Models that include consideration of the organism's life cycle, molecular mechanisms of gene control, and population dynamics are used to describe mutant and wild-type populations in two environments with different demands for expression of the genes in question. These models are analyzed mathematically to identify conditions that lead to either selection or loss of a given mode of control. It will be shown that this theory ties together a number of important variables, including growth rates, mutation rates, minimum and maximum demands for gene expression, and minimum and maximum durations for the life cycle of the organism. An application of the theory is provided in the accompanying article (SAVAGEAU 1998), where regulation of the *lac* and *mal* operons of *E. coli* is analyzed and the results are compared with independent experimental data.

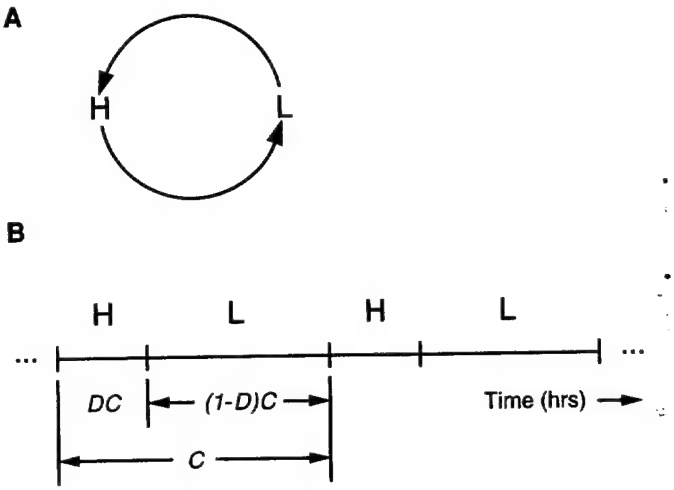


FIGURE 1.—The life cycle of an organism alternating between two different environments. (A) Expression of the genes that are specifically required for growth in the environment labeled H is in high demand, whereas in the alternative environment labeled L their expression is in low demand. (B) The average time required for the organism to complete its life cycle is denoted by *C*. The fraction of its cycle time spent in environment H is denoted by *D*, which also represents demand for expression of the H-specific genes.

MODELS

Life cycle: We shall consider a given effector gene in an organism that cycles between two alternative environments, a high-demand environment H, and a low-demand environment L, as shown in Figure 1. The average *cycle time* required for one complete passage through both H and L environments is denoted by *C*. The average fraction of time spent in the high-demand environment is denoted by *D*. Note that *D* also signifies *demand for expression* of the regulated effector gene. If $D = 0$, demand is minimal because the organism is always in the low-demand environment; if $D = 1$, demand is maximal because the organism is always in the high-demand environment.

Gene expression: The models of gene expression and mutation that will be treated are shown schematically in Figures 2 and 3. The effector genes in each case are normally expressed in environment H but not in environment L. To simplify the diagrams and the discussion, we shall consider mutations in the regulatory mechanism to be an alteration in the modulator site. Mutations in the structural gene for the regulator protein also can disrupt the normal interaction between the regulatory protein and the modulator site to which it binds, and these will be suitably accounted for even though they will not be represented diagrammatically or discussed in detail. Other types of mutations will be considered briefly in the DISCUSSION section.

In the negative mode of control (Figure 2), environment H involves expression of the effector gene in the wild-type organism. It also involves expression in the mutants with a defect in the modulator site to which the negative regulator binds. Normal expression is prevented in the mutants with a defect in the promoter site. Environment L involves the absence of expression of the effector gene in the wild-type organism and in the mutants with a defect in the promoter site. There is inappropriate expression in the mutant with a defect in the modulator site. The mutation rates between the different populations are as indicated.

In the positive mode of control (Figure 3), environ-

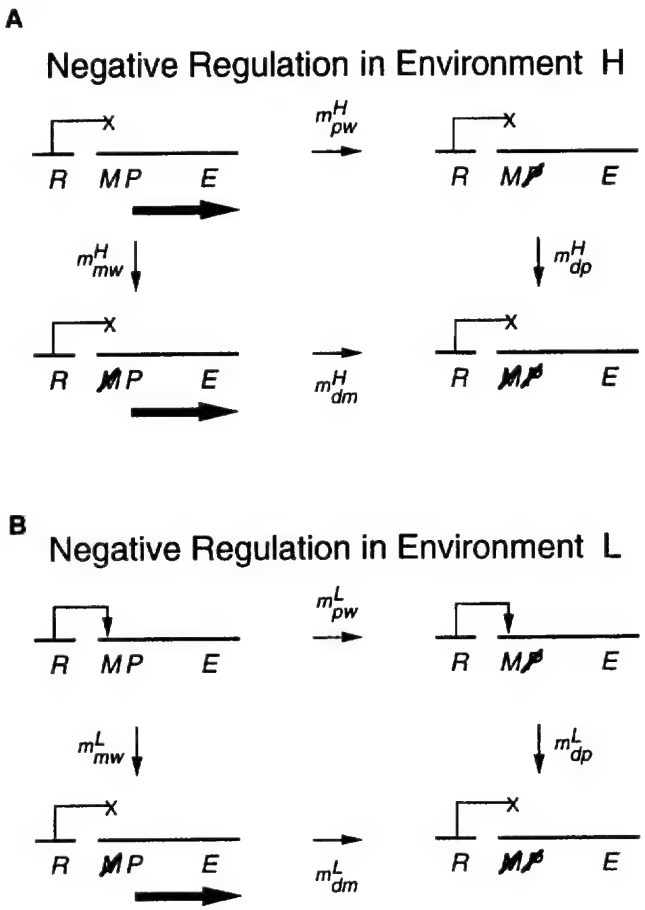


FIGURE 2.—Expression of genes governed by the negative mode of control in the high-demand (H) and low-demand (L) environments. The symbols are as follows: structural gene for the regulator protein, *R*; structural gene for the effector protein, *E*; nucleotide sequence for the promoter site, *P*; and nucleotide sequence for the modulator site, *M*. The wild-type promoter in the negative mode must be a high-level promoter to achieve full expression upon removal of repressor, and a functional modulator site (operator) is necessary for expression to be turned off in the presence of repressor. The heavy arrows indicate transcription of the effector gene. The four diagrams in A and B represent the genotypes of the wild-type (w), promoter mutant (p), modulator mutant (m), and double mutant (d). The mutation rates between the populations of organisms that harbor each of these genotypes are as indicated with the appropriate subscripts and superscripts; *e.g.*, m^{H}_{dm} represents the mutation rate in the high-demand environment for production of double mutants (d) from modulator mutants (m).

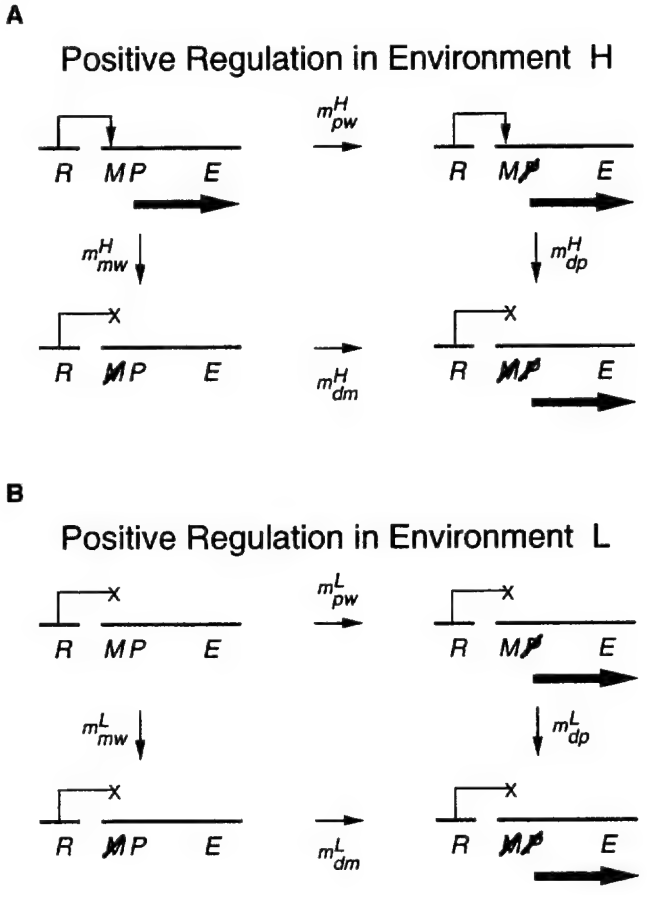


FIGURE 3.—Expression of genes governed by the positive mode of control in the high-demand (H) and low-demand (L) environments. The symbols are as follows: structural gene for the regulator protein, *R*; structural gene for the effector protein, *E*; nucleotide sequence for the promoter site, *P*; and nucleotide sequence for the modulator site, *M*. The wild-type promoter in the positive mode must be a low-level promoter for expression to be turned off upon removal of activator, and a functional modulator site (initiator) is necessary to achieve full expression in the presence of activator. The heavy arrows indicate transcription of the effector gene. The four diagrams in A and B represent the genotypes of the wild-type (w), promoter mutant (p), modulator mutant (m), and double mutant (d). The mutation rates between the populations of organisms that harbor each of these genotypes are as indicated with the appropriate subscripts and superscripts; *e.g.*, m^{L}_{pw} represents the mutation rate in the low-demand environment for production of promoter mutants (p) from wild-type organisms (w).

ment H involves expression of the effector gene in the wild-type organism. It also involves expression in the mutants with a mutationally enhanced promoter site. Normal expression is prevented in the mutants with a defect in the modulator site. Environment L involves the absence of expression of the effector gene in the wild-type organism and in the mutants with a defect in the modulator site. There is inappropriate expression in the mutants with a mutationally enhanced promoter site. The mutation rates between the different populations are as indicated, but it should be noted that the values for these parameters need not be the same for the two modes of control.

Populations: All of the relevant populations and conditions can be represented in a common abstract diagram in which the growth rates of the individual populations and the mutation rates between populations are explicitly depicted (Figure 4). There will be four sets of parameter values associated with this diagram, one each for the negative mode in high demand, the nega-

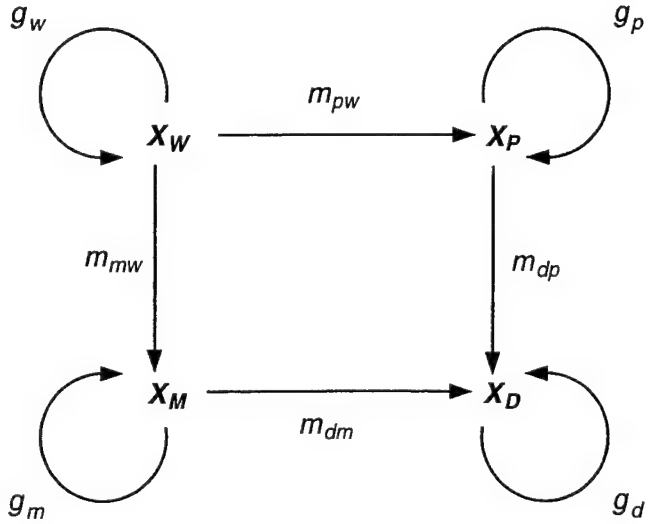


FIGURE 4.—Schematic diagram representing the populations of wild-type and mutant organisms. The symbols are as follows: number of wild-type organisms, X_W; number of promoter mutants, X_P; number of modulator mutants, X_M; and number of double mutants, X_D. The growth rates of each population are indicated by the symbol g with the relevant subscripts, and the mutation rates between populations are indicated by m with the appropriate subscripts. See text for further discussion.

tive mode in low demand, the positive mode in high demand, and the positive mode in low demand.

Assumptions: These models are based on a number of assumptions. First, the organisms harboring these gene systems are assumed to be otherwise isogenic. Second, because we are interested in the conditions for selection of the wild-type regulatory mechanism, we shall assume that the ratio of wild-type to mutant organisms is initially 1/10 its steady-state value and then examine the conditions that lead to enrichment of the wild type. Third, sites in the DNA consist of a number of critical bases, and mutation in any one of these leads to a loss of function in the modulator sites. The same is true of the high-level promoter in the negative mode. The low-level promoter in the positive mode consists of a smaller number of critical bases, and mutation in any of these leads to a mutationally enhanced promoter level. Fourth, the regulator gene consists of a number of critical bases, and mutation in any one of these leads to a loss of the regulator function. Fifth, we will be concerned only with the forward mutational events as indicated in Figures 2–4. The back mutational events can be neglected because the mutant populations will be small, according to our criterion for selection, and the probability of back mutation is lower than that in the forward direction. Sixth, although our models will account for the dynamics of the doubly mutant population, we will neglect this aspect because the singly mutant populations will be small and the probability of a second mutation will make the production rate of the doubly mutant population that much smaller. Finally, we shall assume that expression is fully ON or fully OFF and that both the positive and negative modes of control have the same capacity for gene regulation (SAVAGEAU 1989), which we take to be 100 for the ratio of full expression to basal expression.

PARAMETERS

The macroscopic parameters in our theory can be decomposed into constituent parameters that are defined in terms of reference values and relative values for mutation rates and growth rates.

Mutation rates: The reference mutation rate μ is given by the spontaneous mutation rate per base per DNA replication. The spontaneous mutation rate for various structures in our model can be determined from estimates of the spontaneous mutation rate per base and the relative mutation rate given by the number of critical bases that define the DNA targets for these structures. We will consider the following relative mutation rates in our model: π for loss of a high-level promoter site, υ for gain of a high-level promoter site, τ for loss of a regulator's functional target site, and ρ for loss of a functional regulator protein. We can also define a relative mutation rate ε and explore the effects of gene expression on mutation rate (DATTA and JINKS-ROBERTSON 1995; FRANCINO *et al.* 1996).

Growth rates: The reference growth rate γ is defined as the growth rate of the wild-type organism in the nutritionally richer of the two environments. Its value is not critical because one can simply rescale time accordingly and none of our results would change. The growth rates in other circumstances can be expressed as the product of the reference growth rate and the appropriate relative growth rate. We will consider the following relative growth rates in our model: λ for mutants that have lost normal expression of the effector gene, σ for mutants that exhibit superfluous expression of the effector gene, and δ for the more nutritionally deficient of the two environments.

Criterion for selection: Our criterion for selection is that each mutant population shall be reduced to no more than θ of the wild-type population. A typical value for θ is 0.05% (LECLERC *et al.* 1996).

These relationships are summarized in Table 1. Numerical estimates for these parameters are given in the accompanying article (SAVAGEAU 1998), which provides a specific application of the theory.

QUANTITATIVE DEVELOPMENT OF THE THEORY

The mathematical analysis needed for this development can be significantly reduced by taking advantage of two fundamental symmetries in our model. First, there is a symmetry between the promoter-mutant and modulator-mutant populations that is evident in Figure 4. If the subscripts p and m are simply interchanged the model remains unchanged. This means that we need

TABLE 1

Decomposition of macroscopic parameters into constituent parameters

Parameter[a]	Mode of control: Negative, High demand	Negative, Low demand	Positive, High demand	Positive, Low demand
g_w	γ	$\gamma\delta$	$\gamma\delta$	γ
g_p	$\gamma\lambda$	$\gamma\delta$	$\gamma\delta$	$\gamma\sigma$
g_m	γ	$\gamma\delta\sigma$	$\gamma\delta\lambda$	γ
g_d	$\gamma\lambda$	$\gamma\delta$	$\gamma\delta$	$\gamma\sigma$
m_{pw}	$\mu\pi\varepsilon$	$\mu\pi$	$\mu\upsilon\varepsilon$	$\mu\upsilon$
m_{mw}	$\mu\ (\tau + \rho)\varepsilon$	$\mu\ (\tau + \rho)$	$\mu\ (\tau + \rho)\ \varepsilon$	$\mu\ (\tau + \rho)$
m_{dp}	$\mu\ (\tau + \rho)$	$\mu\ (\tau + \rho)$	$\mu\ (\tau + \rho)\ \varepsilon$	$\mu\ (\tau + \rho)\ \varepsilon$
m_{dm}	$\mu\pi\varepsilon$	$\mu\pi\varepsilon$	$\mu\upsilon$	$\mu\upsilon$

See Figures 2–4 for definition of parameters.

[a] The parameters for growth rates and mutation rates in turn determine the parameters for the rate constants in the dynamic Equations 1–4: $\alpha_{ww} = [1 - (m_{pw} + m_{mw})]g_w$, $\alpha_{pw} = m_{pw}g_w$, $\alpha_{pp} = (1 - m_{dp})g_p$, $\alpha_{mw} = m_{mw}g_w$, $\alpha_{mm} = (1 - m_{dm})g_m$, $\alpha_{dm} = m_{dm}g_m$, $\alpha_{dp} = m_{dp}g_p$, $\alpha_{dd} = g_d$.

only carry out the analysis for the promoter-mutant population; the corresponding results for the modulator-mutant population can then be obtained simply by interchanging the subscripts p and m. Second, there is a symmetry between the first and second phases of the cycle depicted in Figure 1. If the H and L phases are interchanged along with the symbols D and $(1 - D)$ the temporal pattern remains unchanged. This means that we need only carry out the analysis from the beginning of the H phase; the corresponding results from the beginning of the L phase can then be obtained by interchanging the superscripts H and L and the symbols D and $(1 - D)$.

Dynamics: The equations describing the dynamic behavior of the model in Figure 4 are

$$dX_w/dt = \alpha_{ww} X_w \tag{1}$$

$$dX_p/dt = \alpha_{pw} X_w + \alpha_{pp} X_p \tag{2}$$

$$dX_m/dt = \alpha_{mw} X_w + \alpha_{mm} X_m \tag{3}$$

$$dX_d/dt = \alpha_{dm} X_m + \alpha_{dp} X_p + \alpha_{dd} X_d, \tag{4}$$

where the numbers for each population as a function of time are given by the symbol X with appropriate subscripts and the first-order rate constants are given by the symbol α, again with appropriate subscripts. The rate constants are in turn related to the various mutation rates and growth rates, represented by the symbols m and g with suitable subscripts: $\alpha_{ww} = [1 - (m_{pw} + m_{mw})]g_w$, $\alpha_{pw} = m_{pw}g_w$, $\alpha_{pp} = (1 - m_{dp})g_p$, $\alpha_{mw} = m_{mw}g_w$, $\alpha_{mm} = (1 - m_{dm})g_m$, $\alpha_{dm} = m_{dm}g_m$, $\alpha_{dp} = m_{dp}g_p$, $\alpha_{dd} = g_d$.

Equations 1–4 are linear and easily solved to obtain numbers for the wild-type and mutant populations as a function of time. The numbers for the wild-type and promoter-mutant populations at the end of a full period in environment H are given in terms of the initial values at an arbitrary time t:

$$X_w(t + DC) = X_w(t) \exp[\alpha^H_{ww} DC] \tag{5}$$

$$\begin{aligned} X_p(t + DC) = {} & [\alpha^H_{pw}/(\alpha^H_{ww} - \alpha^H_{pp})] X_w(t) \exp[\alpha^H_{ww} DC] \\ & + \{X_p(t) - [\alpha^H_{pw}/(\alpha^H_{ww} - \alpha^H_{pp})] X_w(t)\} \\ & \times \exp[\alpha^H_{pp} DC]. \end{aligned} \tag{6}$$

These numbers then become the initial values for the solution in environment L, and the numbers at the end of the period in environment L are then

$$X_w(t + C) = X_w(t) \exp[\alpha^H_{ww} DC] \exp[\alpha^L_{ww}(1 - D)C] \tag{7}$$

$$\begin{aligned} X_p(t + C) = {} & X_w(t) \{[\alpha^L_{pw}/(\alpha^L_{ww} - \alpha^L_{pp})] \exp[\alpha^H_{ww} DC] \\ & \times \{\exp[\alpha^L_{ww}(1 - D)C] - \exp[\alpha^L_{pp}(1 - D)C]\} \\ & + [\alpha^H_{pw}/(\alpha^H_{ww} - \alpha^H_{pp})] \exp[\alpha^L_{pp}(1 - D)C] \\ & \times \{\exp[\alpha^H_{ww} DC] - \exp[\alpha^H_{pp} DC]\}\} \\ & + X_p(t) \exp[\alpha^H_{pp} DC] \exp[\alpha^L_{pp}(1 - D)C]. \end{aligned} \tag{8}$$

Thus, the temporal behavior is determined by four exponential functions with time constants that are independent of C.

The ratio of the promoter-mutant to the wild-type numbers, which is plotted in Figure 5, yields

$$\begin{aligned} X_p(t + C)/X_w(t + C) = {} & \{[\alpha^L_{pw}/(\alpha^L_{ww} - \alpha^L_{pp})] \\ & \times \{1 - \exp[(\alpha^L_{pp} - \alpha^L_{ww})(1 - D)C]\} \\ & + [\alpha^H_{pw}/(\alpha^H_{ww} - \alpha^H_{pp})] \\ & \times \{1 - \exp[(\alpha^H_{pp} - \alpha^H_{ww})DC]\} \\ & \times \exp[(\alpha^L_{pp} - \alpha^L_{ww})\ (1 - D)C]\} \\ & + \{\exp[(\alpha^H_{pp} - \alpha^H_{ww})DC] \\ & \times \exp[(\alpha^L_{pp} - \alpha^L_{ww})(1 - D)C]\} \\ & \times X_p(t)/X_w(t) \end{aligned}$$

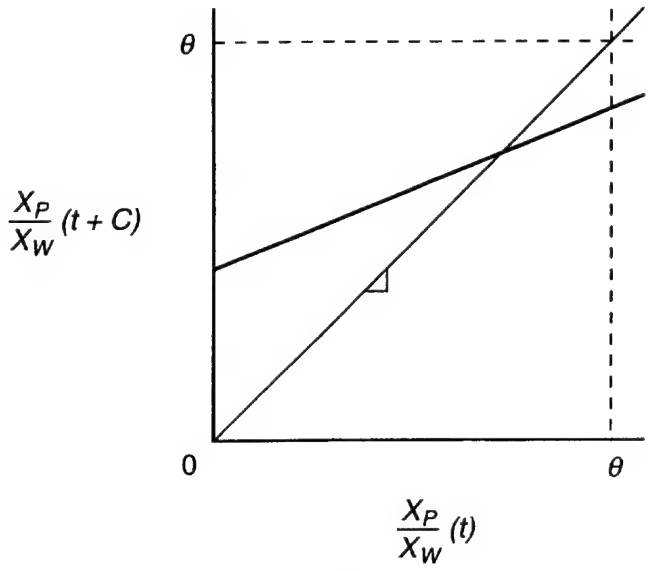


FIGURE 5.—Recursive relationship for the ratio of population sizes for promoter-mutant and wild-type organisms. The horizontal axis gives the value of the ratio at an arbitrary time t; the vertical axis gives the value at the subsequent time $t + C$, which is one complete cycle later. Selection for the wild-type organism is indicated when the recursive relationship, which is the straight line given by Equation 9, has a slope between 0 and 1 and an intercept between 0 and 0.0005. The intersection of this line with the 45° line determines a value for the ratio that represents a stable steady state.

or

$$X_p(t + C)/X_w(t + C) = (\text{intercept}) + (\text{slope})\ X_p(t)/X_w(t). \quad (9)$$

Note that the intercept and slope in this expression are both positive quantities. A slope greater than 1 implies that the ratio tends to infinity with time and thus that the wild-type promoter is lost. A slope between 0 and 1 implies that the ratio tends to a fixed value (given by the intersection with the 45° line) with time and, if this value is less than θ (the criterion for selection), that the wild-type promoter will be preserved. An intercept greater than θ implies loss of the wild-type promoter no matter what the value of the slope.

Starting with any set of values for the wild-type and promoter-mutant populations, Equations 7–9 can be applied recursively to calculate the subsequent population sizes and ratios as a function of time. From these results one can determine the rate of selection of the wild-type regulatory mechanism.

Steady-state pattern: The ratio of promoter-mutant and wild-type populations increases in one environment and decreases in the other to produce a sawtooth pattern. Once the initial transients have died away, a repeating pattern with two steady-state values is established. The first value of the ratio in steady state, when it exists, is calculated by equating the ratios on the two sides of Equation 9 and solving to obtain the following expression:

$$\begin{aligned} X_p/X_w = {} & \{[\alpha^L_{pw}/(\alpha^L_{ww} - \alpha^L_{pp})] \\ & \times \{1 - \exp[(\alpha^L_{pp} - \alpha^L_{ww})(1 - D)C]\} \\ & + [\alpha^H_{pw}/(\alpha^H_{ww} - \alpha^H_{pp})] \\ & \times \{1 - \exp[(\alpha^H_{pp} - \alpha^H_{ww})DC]\} \exp[(\alpha^L_{pp} - \alpha^L_{ww}) \\ & \times (1 - D)C]\}/\{1 - \exp[(\alpha^H_{pp} - \alpha^H_{ww})DC \\ & + (\alpha^L_{pp} - \alpha^L_{ww})(1 - D)C]\}. \end{aligned} \quad (10)$$

If, instead of starting the analysis at the beginning of the period in environment H, we were to start it at the beginning of the period in environment L, then the results would be equivalent to those in Equations 5–10 except for an exchange of the superscripts H and L and the symbols D and $(1 - D)$. The second value of the ratio in steady state, when it exists, is thus

$$\begin{aligned} X_p/X_w = {} & \{[\alpha^H_{pw}/(\alpha^H_{ww} - \alpha^H_{pp})]\{1 - \exp[(\alpha^H_{pp} - \alpha^H_{ww})DC]\} \\ & + [\alpha^L_{pw}/(\alpha^L_{ww} - \alpha^L_{pp})]\{1 - \exp[(\alpha^L_{pp} - \alpha^L_{ww}) \\ & \times (1 - D)C]\} \\ & \times \exp[(\alpha^H_{pp} - \alpha^H_{ww})DC]\}/\{1 - \exp[(\alpha^L_{pp} - \alpha^L_{ww}) \\ & \times (1 - D)C + (\alpha^H_{pp} - \alpha^H_{ww})DC]\}. \end{aligned} \quad (11)$$

Equations 10 and 11 represent different aspects of the same steady-state pattern. One of the two steady-state solutions for this ratio gives the maximum value whereas the other gives the minimum value. These values can be used to define the extent of selection. We shall always be interested in the maximum value of the ratio; if this is less than the criterion for selection, then the minimum value will certainly be less as well.

Definition of the threshold for selection: The threshold for selection of the wild-type promoter is obtained from the solution of Equation 10 or 11, whichever gives the maximum value for the ratio. The values for the growth rates and mutation rates in the high- and low-demand environments (for either the positive or the negative mode of control in Table 1) determine the values for the rate-constant parameters that appear in Equations 10 and 11. The ratio X_p/X_w is then fixed with a value equal to θ, which is the criterion for selection. The result of these parameter assignments is a nonlinear equation involving the cycle time C and the demand for gene expression D that defines the *threshold for selection*. There is no explicit solution for C as a function of D. However, the threshold for selection of the wild-type promoter can be obtained by bisection (PRESS *et al.* 1988) when numerical values are assumed for the parameters in Equation 10 or 11.

As noted at the beginning of this section, the corresponding results for the modulator-mutant population can be obtained from Equations 5–11 simply by inter-

changing the subscripts p and m. We will make use of these expressions below.

Although there is no analytical solution that gives the thresholds for selection, their asymptotic behavior can be determined analytically. As will be seen in the following sections, the analytical expressions allow one to draw general conclusions that are independent of particular numerical values for the parameters.

Threshold for selection of a promoter with the negative mode: The ratio of promoter-mutant and wild-type populations is decreasing in environment H and increasing in environment L. Thus, the maximum value in steady state is determined from the analysis that starts in H. The asymptotic character of the threshold for selection of the promoter can be determined from Equation 10. First, it should be noted from Table 1 that $(\alpha^{L}_{pp} - \alpha^{L}_{ww}) > 0$ and $\alpha^{L}_{pw}/(\alpha^{L}_{ww} - \alpha^{L}_{pp}) = -1$. Second, for typical values of the parameters, $(\alpha^{H}_{pp} - \alpha^{H}_{ww}) < 0$.

When $C \gg 1$, and $D > (\alpha^{L}_{pp} - \alpha^{L}_{ww})/[(\alpha^{L}_{pp} - \alpha^{L}_{ww}) - (\alpha^{H}_{pp} - \alpha^{H}_{ww})]$, Equation 10 can be approximated as

$$\theta = \exp[(\alpha^{L}_{pp} - \alpha^{L}_{ww})(1 - D)C] - 1 + [\alpha^{H}_{pw}/(\alpha^{H}_{ww} - \alpha^{H}_{pp})] \exp[(\alpha^{L}_{pp} - \alpha^{L}_{ww})(1 - D)C], \quad (12)$$

where θ is the criterion for selection of the promoter. Solving for C as a function of D yields

$$C = \frac{\log[1 + \theta] - \log[1 + \alpha^{H}_{pw}/(\alpha^{H}_{ww} - \alpha^{H}_{pp})]}{(\alpha^{L}_{pp} - \alpha^{L}_{ww})} \frac{1}{1 - D}. \quad (13)$$

The arguments of the logarithms are nearly unity, so that

$$C = \frac{\theta - \alpha^{H}_{pw}/(\alpha^{H}_{ww} - \alpha^{H}_{pp})}{(\alpha^{L}_{pp} - \alpha^{L}_{ww})} \frac{1}{1 - D} \quad (14)$$

or

$$C = \frac{\theta - \mu\pi\varepsilon/[1 - \lambda(1 - \mu\tau - \mu\rho) - \mu\varepsilon(\pi + \tau + \rho)]}{\mu\pi\gamma\delta} \times \frac{1}{1 - D} \approx \frac{\theta}{\mu\pi\gamma\delta} \frac{1}{1 - D}. \quad (15)$$

Thus, the high-C asymptote in a log C *vs.* log D plot is given by a line that is nearly horizontal for values of $D \ll 1$ and that approaches infinity as D goes to unity.

When $C \ll 1$, the exponential functions in Equation 10 can be approximated by the first three terms of their Taylor series and the resulting equation can be solved for C as a function of D. The value of $D = D_{min}$ that makes $C = 0$ is given by

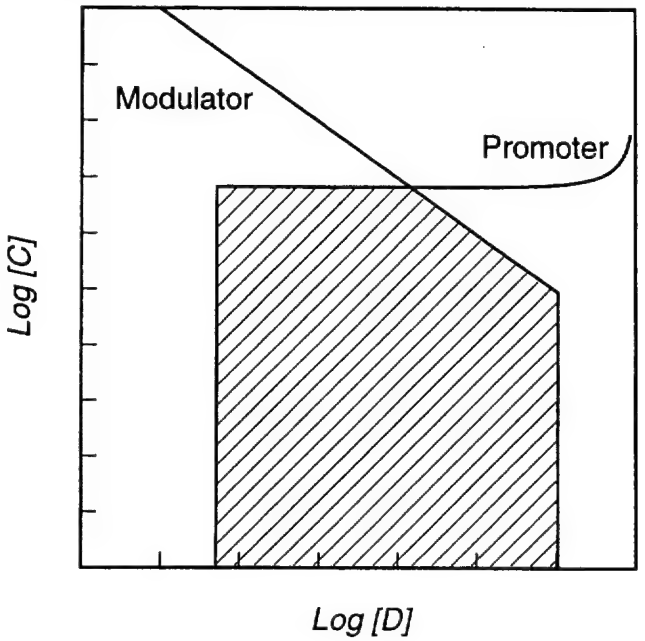


FIGURE 6.—Schematic representation of the thresholds for selection of the wild-type regulatory mechanism as functions of the cycle time and the demand for gene expression. The threshold for selection against the promoter mutants is obtained for a given set of parameter values by setting the ratio of $X_p/X_w = 0.0005$ in Equation 10 or 11 and then solving for the cycle time C as a function of the demand for gene expression D. The threshold for selection against the modulator mutants is obtained in a similar fashion (see text for discussion). In each case, selection is indicated by values for C and D that lie below the calculated threshold. Selection for the wild-type regulatory mechanism occurs for those values of C and D that lie below both threshold simultaneously. These thresholds define minimum and maximum values for demand.

$$D_{min} = \frac{(\alpha^{L}_{pp} - \alpha^{L}_{ww})(1 + \theta)}{(\alpha^{L}_{pp} - \alpha^{L}_{ww})(1 + \theta) - \alpha^{H}_{pw} - (\alpha^{H}_{pp} - \alpha^{H}_{ww})\theta} \quad (16)$$

or

$$D_{min} = \mu\pi\delta(1 + \theta)/\{\mu\pi\delta(1 + \theta) + [1 - \lambda(1 - \mu\tau - \mu\rho) - \mu\varepsilon(\pi + \tau + \rho)]\theta - \mu\pi\varepsilon\} \approx \mu\pi\delta/\theta(1 - \lambda). \quad (17)$$

Thus, the low-C asymptote is given by a vertical line located at $D = D_{min}$ in a log C *vs.* log D plot.

The threshold for selection of the promoter is characterized by the combination of these high- and low-C asymptotes as shown schematically in Figure 6.

Threshold for selection of a modulator (regulator) with the negative mode: The ratio of modulator-mutant and wild-type populations is decreasing in environment L and increasing in environment H. Thus, the maximum value in steady state is determined from the analysis that starts in L. The asymptotic character of the

threshold for selection of the modulator (regulator) can be determined from Equation 11 after interchanging the subscripts p and m. In this case, $(\alpha^H_{mm} - \alpha^H_{ww}) > 0$, $\alpha^H_{mw}/(\alpha^H_{ww} - \alpha^H_{mm}) = -1$ and, for typical values of the parameters, $(\alpha^L_{mm} - \alpha^L_{ww}) < 0$.

When $C \gg 1$, and $D < (\alpha^L_{mm} - \alpha^L_{ww})/[(\alpha^L_{mm} - \alpha^L_{ww}) - (\alpha^H_{mm} - \alpha^H_{ww})]$, Equation 11 can be approximated as

$$\theta = \exp[(\alpha^H_{mm} - \alpha^H_{ww})DC] - 1 + [\alpha^L_{mw}/(\alpha^L_{ww} - \alpha^L_{mm})] \exp[(\alpha^H_{mm} - \alpha^H_{ww})DC], \quad (18)$$

where θ is the criterion for selection of the modulator. Solving for C as a function of D yields

$$C = \frac{\log[1 + \theta] - \log[1 + \alpha^L_{mw}/(\alpha^L_{ww} - \alpha^L_{mm})]}{(\alpha^H_{mm} - \alpha^H_{ww})} \frac{1}{D}. \quad (19)$$

The arguments of the logarithms are nearly unity, so that

$$C = \frac{\theta - \alpha^L_{mw}/(\alpha^L_{ww} - \alpha^L_{mm})}{(\alpha^H_{mm} - \alpha^H_{ww})} \frac{1}{D} \quad (20)$$

or

$$C = \frac{\theta - \mu(\tau + \rho)/[1 - \sigma(1 - \mu\pi\varepsilon) - \mu(\pi + \tau + \rho)]}{\mu(\tau + \rho)\varepsilon\gamma} \frac{1}{D} \approx \frac{\theta}{\mu(\tau + \rho)\varepsilon\gamma} \frac{1}{D}. \quad (21)$$

Thus, the high-C asymptote is given by a straight line with slope equal to -1 in a log C *vs.* log D plot.

When $C \ll 1$, the exponential functions in the steady-state ratio can be approximated by the first three terms of their Taylor series and the resulting equation can be solved for C as a function of D. The value of $D = D_{max}$ that makes $C = 0$ is given by

$$D_{max} = \frac{[-\theta(\alpha^L_{mm} - \alpha^L_{ww}) - \alpha^L_{mw}]}{[-\theta(\alpha^L_{mm} - \alpha^L_{ww}) - \alpha^L_{mw}] + (\alpha^H_{mm} - \alpha^H_{ww})(1 + \theta)} \quad (22)$$

or

$$D_{max} = \delta\{\theta[1 - \sigma(1 - \mu\pi\varepsilon) - \mu(\pi + \tau + \rho)] - \mu(\tau + \rho)\}/\delta\{\theta[1 - \sigma(1 - \mu\pi\varepsilon) - \mu(\pi + \tau + \rho)] - \mu(\tau + \rho)\} + \mu(\tau + \rho)\varepsilon(1 + \theta) \approx 1/\{1 + \mu(\tau + \rho)\varepsilon/[\delta\theta(1 - \sigma)]\}. \quad (23)$$

Thus, the low-C asymptote is given by a vertical line located at $D = D_{max}$ in a log C *vs.* log D plot.

The threshold for selection of the modulator (regulator) is characterized by the combination of these high- and low-C asymptotes as shown schematically in Figure 6.

Region in which selection for the negative mode of control is realizable: Selection for both wild-type promoter and wild-type modulator (regulator) requires values of C and D that lie in the shaded region below the two thresholds shown schematically in Figure 6. The low-C asymptotes of these thresholds (Equations 17 and 23) define the minimum D_{min} and maximum D_{max} values of the demand for gene expression. The intersection of the two thresholds yields a prediction for maximum cycle time C_{max}. As shown elsewhere, with numerical estimates for the various parameters, the theory predicts other more relevant values not only for maximum cycle time, but also for minimum cycle time and optimal cycle time (SAVAGEAU 1998). Thus, the thresholds define a region of the C *vs.* D plot within which selection for the wild-type regulatory mechanism is realizable and outside of which it is not.

Existence of a region of realizable selection for the negative mode: Clearly, $D_{max} > D_{min}$ is required for a region of realizable selection to exist. These boundaries for selection are strongly influenced by the selection coefficients ($1 - \lambda$ and $1 - \sigma$), which are related to the differences in growth rates for wild-type and mutant organisms. This is seen most clearly for the simplified case in which all relative mutation rates are equal to unity and all mutants have the same reduction in growth rate. The inequality involving Equations 17 and 23 yields a critical value for the selection coefficients; selection of the wild-type regulatory mechanism is possible only when the selection coefficients exceed this critical value:

$$(1 - \lambda) = (1 - \sigma) > \frac{\mu(1 + \delta)}{2\theta}\left[1 + \sqrt{1 + \frac{4(1 - \delta)}{(1 + \delta)^2}}\right]. \quad (24)$$

This can be seen graphically in Figure 7 where the thresholds for selection are plotted for different values of the selection coefficients.

Discriminate selection for the negative mode of control: When the reduction in growth rate for the mutants is sufficiently small ($<\sim$0.0005% in this illustration) there is no overlap beneath the thresholds. No selection for the wild-type regulatory mechanism is possible when the selection pressure is too weak. When the reduction in growth rate has an intermediate value (between 0.0005 and 0.01% in this illustration) there is a significant and well-delineated overlap beneath the thresholds. Discriminate selection for the wild-type regulatory mechanism occurs within a range of relatively *low* values for demand, but not outside it. When the reduction in growth rate is sufficiently large ($>\sim$0.01% in this illustration) the overlap is so large that it encompasses almost the entire range of values for demand. Indiscriminate selection for the wild-type regulatory mechanism occurs under these conditions.

Threshold for selection of a promoter with the positive mode: The ratio of promoter-mutant and wild-type

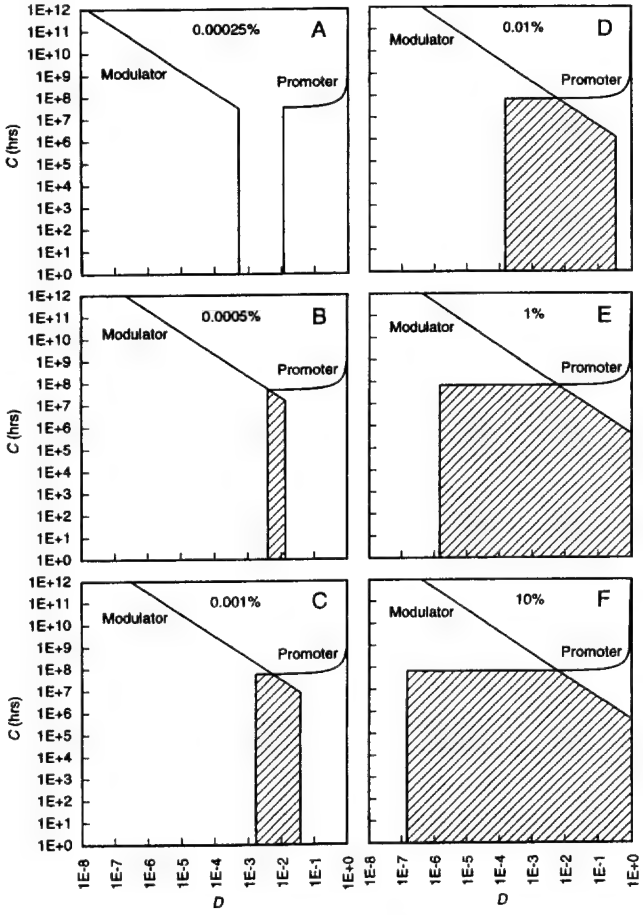


FIGURE 7.—Discriminate selection for wild-type regulatory mechanisms with alternative modes of control requires intermediate values for the selection coefficients. Results (A–F) are shown for the negative mode in a simplified case (see text for discussion). When selection coefficients are too low (<0.0005%), there is no selection for the wild type. At intermediate values (0.0005–0.01%), discriminate selection for the wild type occurs at relatively low values of demand. When selection coefficients are too high (>0.01%), selection for the wild-type regulatory mechanism occurs indiscriminately at nearly all values of demand. The results for the positive mode are similar, except that discriminate selection occurs at relatively high values of demand.

populations is decreasing in environment L and increasing in environment H. Thus, the maximum value in steady state is determined from the analysis that starts in L. The asymptotic character of the threshold for selection of the promoter can be determined from Equation 11. In this case, it can be seen from Table 1 that $(\alpha^H_{pp} - \alpha^H_{ww}) > 0$, $\alpha^H_{pw}/(\alpha^H_{ww} - \alpha^H_{pp}) = -1$ and, for typical values of the parameters, $(\alpha^L_{pp} - \alpha^L_{ww}) < 0$.

When $C \gg 1$ and $(1 - D) > (\alpha^H_{pp} - \alpha^H_{ww})/[(\alpha^H_{pp} - \alpha^H_{ww}) - (\alpha^L_{pp} - \alpha^L_{ww})]$, Equation 11 can be approximated as

$$\theta = \exp[(\alpha^H_{pp} - \alpha^H_{ww})DC] - 1 + [\alpha^L_{pw}/(\alpha^L_{ww} - \alpha^L_{pp})] \exp[(\alpha^H_{pp} - \alpha^H_{ww})DC]. \quad (25)$$

Solving for C as a function of $1 - D$ yields

$$C = \frac{\log[1 + \theta] - \log[1 + \alpha^L_{pw}/(\alpha^L_{ww} - \alpha^L_{pp})]}{(\alpha^H_{pp} - \alpha^H_{ww})} \times \frac{1}{1 - (1 - D)} \quad (26)$$

or

$$C = \frac{\theta - \mu\upsilon/[1 - \sigma(1 - \mu(\tau + \rho)\varepsilon) - \mu(\upsilon + \tau + \rho)]}{\mu\upsilon\varepsilon\gamma\delta} \times \frac{1}{1 - (1 - D)} \approx \frac{\theta}{\mu\upsilon\varepsilon\gamma\delta} \frac{1}{1 - (1 - D)}. \quad (27)$$

Thus, the high-C asymptote in a log C *vs.* $\log(1 - D)$ plot is given by a line that is nearly horizontal for values of $(1 - D) \ll 1$ and that approaches infinity as $(1 - D)$ goes to unity.

When $C \ll 1$, the exponential functions in Equation 11 can be approximated by the first three terms of their Taylor series and the resulting equation can be solved for C as a function of $1 - D$. The value of $1 - D = 1 - D_{max}$ that makes $C = 0$ is given by

$$1 - D_{max} = \frac{(\alpha^H_{pp} - \alpha^H_{ww})(1 + \theta)}{(\alpha^H_{pp} - \alpha^H_{ww})(1 + \theta) - \alpha^L_{pw} - (\alpha^L_{pp} - \alpha^L_{ww})\theta} \quad (28)$$

or

$$1 - D_{max} = \mu\upsilon\varepsilon\delta(1 + \theta)/\{\mu\upsilon\varepsilon\delta(1 + \theta) + [1 - \sigma(1 - \mu(\tau + \rho)\varepsilon) - \mu(\upsilon + \tau + \rho)]\theta - \mu\upsilon\} \approx \mu\upsilon\varepsilon\delta/\theta(1 - \sigma) . \quad (29)$$

Thus, the low-C asymptote is given by a vertical line located at $1 - D = 1 - D_{max}$ in a log C *vs.* $\log(1 - D)$ plot.

The threshold for selection of the promoter in this case is characterized by high- and low-C asymptotes that are similar to those for the negative mode shown schematically in Figure 6, except that the horizontal axis is given by $\log(1 - D)$ rather than $\log D$ (data not shown).

Threshold for selection of a modulator (regulator) with the positive mode: The ratio of modulator-mutant and wild-type populations is decreasing in environment H and increasing in environment L. Thus, the maximum value in steady state is determined from the analysis that starts in H. The asymptotic character of this threshold can be determined from Equation 10 after interchanging the subscripts p and m. In this case, $(\alpha^L_{mm} - \alpha^L_{ww}) > 0$, $\alpha^L_{mw}/(\alpha^L_{ww} - \alpha^L_{mm}) = -1$ and, for typical values of the parameters, $(\alpha^H_{mm} - \alpha^H_{ww}) < 0$.

When $C \gg 1$, and $(1 - D) < (\alpha^H_{mm} - \alpha^H_{ww})/[(\alpha^H_{mm} -$

$\alpha^H_{ww}) - (\alpha^L_{mm} - \alpha^L_{ww})]$, Equation 10 can be approximated as

$$\theta = \exp[(\alpha^L_{mm} - \alpha^L_{ww})(1 - D)C] - 1 + [\alpha^H_{mw}/(\alpha^H_{ww} - \alpha^H_{mm})] \exp[(\alpha^L_{mm} - \alpha^L_{ww})(1 - D)C] . \quad (30)$$

Solving for C as a function of $1 - D$ yields

$$C = \frac{\log[1 + \theta] - \log[1 + \alpha^H_{mw}/(\alpha^H_{ww} - \alpha^H_{mm})]}{(\alpha^L_{mm} - \alpha^L_{ww})} \frac{1}{1 - D} \quad (31)$$

or

$$C = \frac{\theta - \mu(\tau + \rho)\varepsilon/[1 - \lambda(1 - \mu\upsilon) - \mu\varepsilon(\upsilon + \tau + \rho)]}{\mu(\tau + \rho)\gamma} \times \frac{1}{1 - D} \quad (32)$$

$$\approx \frac{\theta}{\mu(\tau + \rho)\gamma} \frac{1}{1 - D} .$$

Thus, the high-C asymptote is given by a straight line with slope equal to -1 in a log C *vs.* $\log(1 - D)$ plot.

When $C \ll 1$, the exponential functions in the steady-state ratio can be approximated by the first three terms of their Taylor series and the resulting equation can be solved for C as a function of $1 - D$. The value of $1 - D = 1 - D_{min}$ that makes $C = 0$ is given by

$$1 - D_{min} = \frac{-\alpha^H_{mw} - (\alpha^H_{mm} - \alpha^H_{ww})\theta}{-\alpha^H_{mw} - (\alpha^H_{mm} - \alpha^H_{ww})\theta + (\alpha^L_{mm} - \alpha^L_{ww})(1 + \theta)} \quad (33)$$

or

$$1 - D_{min} = \frac{\delta\{\theta[1 - \lambda(1 - \mu\upsilon) - \mu\varepsilon(\upsilon + \tau + \rho)] - \mu(\tau + \rho)\varepsilon\}}{\delta\{\theta[1-\lambda(1-\mu\upsilon)-\mu\varepsilon(\upsilon+\tau+\rho)]-\mu(\tau+\rho)\varepsilon\}+\mu(\tau+\rho)(1+\theta)}$$

$$\approx \frac{1}{1 + \mu(\tau + \rho)/[\delta\theta(1 - \lambda)]} . \quad (34)$$

Thus, the low-C asymptote is given by a vertical line located at $1 - D = 1 - D_{max}$ in a log C *vs.* $\log(1 - D)$ plot.

The threshold for selection of the modulator (regulator) in this case is characterized by high- and low-C asymptotes that are similar to those for the negative mode shown schematically in Figure 6, except that the horizontal axis is given by $\log(1 - D)$ rather than $\log D$ (data not shown).

Discriminate selection for the positive mode of control: The results for the positive mode of control are completely symmetrical to those obtained for the negative mode of control under the simplifying conditions in Figure 7; one need only replace D by $(1 - D)$. When the percentage reduction in growth rate for the mutants is small, no selection for the wild-type regulatory mechanism is possible. At intermediate percentages, discriminate selection for the positive mode of control occurs within a well-delineated range of relatively *high* values for demand, but not outside this range. At large percentages, selection occurs indiscriminately at nearly all values for demand, and, given the above results for the negative mode, one would expect positive and negative modes of control to arise at random with nearly equal probability. Such indiscriminate selection is inconsistent with the experimental evidence, which suggests discriminate selection of negative and positive modes of control based on demand for gene expression (Savageau 1989).

Asymmetric regions in which selection for the alternative modes is realizable: The simplified case examined in Figure 7 suggests completely symmetric regions in which selection for the alternative modes occurs. Alternatively, the region for the positive mode with $1 - D$ as the horizontal axis is identical to that for the negative mode with D as the horizontal axis. This implies that the value of D_{max} (Equation 23) for the negative mode is equal to the value of $1 - D_{min}$ (Equation 34) for the positive mode. This would be true if the following conditions were satisfied: $\theta_N = \theta_P$, $\mu_N = \mu_P$, $\tau_N = \tau_P$, $\rho_N = \rho_P$, $\varepsilon_N = \varepsilon_P = 1$, $\sigma_N = \lambda_P$, $\pi_N = \upsilon_P$. While it is reasonable to assume that the first four conditions are satisfied (criterion for selection θ, mutation rate μ, size of the modulator target τ, and size of the regulator ρ are the same for both the negative N and positive P mode), it is very unlikely that the last three would ever be satisfied. There is evidence that gene expression has an influence on mutation rate ($\varepsilon \neq 1$), that the reduction in growth rate due to superfluous gene expression is less than that due to the loss of normal gene expression ($\sigma_N < \lambda_P$), and that down-promoter mutations in the negative mode are more frequent than up-promoter mutations in the positive mode ($\pi_N > \upsilon_P$). From these considerations we can predict asymmetric regions in which selection for the alternative modes is realizable. Furthermore, because loss of normal expression typically causes a more significant reduction in growth rate than superfluous expression, we can predict that the realizable region for selection of the positive mode is greater than that for the negative mode.

Time course of selection: If we start with each mutant ratio (X_p/X_w and X_m/X_w) at some value larger than its steady-state value, then these mutant ratios will monotonically decrease with time, as can be seen from Figure 5. Alternatively, the wild-type regulatory mechanism is enriched with time, since the ratio of wild-type to mutant organisms $X_w/(X_m + X_p)$ is equal to the reciprocal of the *mutant fraction*, which we define as f_m. The temporal behavior of the populations is a function of the demand for gene expression D. However, the behavior is independent of the cycle time C in the following sense. The time scale is actually discrete, given by values of nC, where n is the number of cycles. Thus, within a fixed time period, the same degree of enrichment can be

achieved with either a large value for C and a small number n or a small value for C and a larger number n.

Extent of selection: While there is selection for the wild-type regulatory mechanism throughout the region of overlap beneath the thresholds (*e.g.*, Figure 6), the extent of the selection varies as a function of cycle time C and demand D. We define the *extent of selection* as the steady-state value of $X_w/(X_m + X_p)$, which is the inverse of the mutant fraction in the population ($1/f_m$). For a given value of $C < C_{max}$, one mutant population increases as the corresponding threshold is approached; it dominates the mutant fraction and the extent of selection reaches its minimum ($1/\theta$). Similarly, the second mutant population increases as the other threshold is approached; it dominates the mutant fraction and the extent of selection again reaches its minimum. Thus, the extent of selection reaches its maximum at a value of D that is intermediate between its threshold values.

Rate of selection: Equations 7–9 can be applied recursively to calculate population sizes and ratios as a function of time. The rate at which selection occurs is independent of cycle time, as noted above. We define *response time* as the time required for the ratio $X_w/(X_m + X_p)$ to reach 99% of its steady-state value starting from an initial state in which the numbers of the two types of mutants are equal and the ratio is equal to 1/10 of its steady-state value. Recall that the time points are given in units of nC, where C is the cycle time and n is the number of cycles. The same temporal behavior is obtained regardless of whether C is large (n small) or small (n large). However, the resolution is poorer for large values of C because the minimum value of n is 1. There is no analytical expression for response time, but it is readily determined by numerical means in specific cases, as can be seen in the following application (SAVAGEAU 1998).

DISCUSSION

Demand theory of gene regulation predicts that the molecular mode of control is correlated with the demand for gene expression in the organism's natural environment (SAVAGEAU 1989). The quantitative development presented in this article not only confirms and quantifies the previous qualitative predictions, but it also identifies critical factors and reveals new relationships.

The recursive equations that characterize the population dynamics of mutant and wild-type organisms (Equations 7–9) allow one to predict the time course for selection. The form of these equations also allows one to predict that the response time for selection is independent of the cycle time C, whereas it is strongly dependent upon the demand for gene expression D. The steady-state solution of the recursive equations provides estimates for the extent of selection (Equations 10 and 11). A threshold for selection is determined by the relationship between cycle time C and demand D that results when the extent of selection is set equal to the criterion for selection.

The thresholds for selection in the C *vs.* D plot define regions within which selection of the positive or negative mode of regulation is realizable (Figure 6). Their intersection defines a maximum value for the cycle time C_{max}, and their asymptotes define minimum D_{min} and maximum D_{max} values of the demand for gene expression. These regions also exhibit an inherent asymmetry that favors selection of the positive mode.

As can be seen from the asymptotic expressions for D_{min} and D_{max} (Equations 17, 23, 29, and 34), the ratio of mutation rate to selection coefficient is the most relevant determinant of the allowed region for selection. Indeed, if the target sizes for the various types of mutations and the selection coefficients are increased by the same order of magnitude, then the results are essentially unchanged.

These predictions, and others that are made possible by the assignment of specific values for the parameters, are examined further in the accompanying article (SAVAGEAU 1998), where we apply this theory to the regulation of the lactose and maltose operons of *Escherichia coli.*

The quantitative version of demand theory presented in this study provides a framework for further development. Other types of mutations can be incorporated in a relatively straightforward manner. Mutations that result in a phenotype similar to that of an existing mutation can be included by simply adding their target size, as was done here for mutations in the regulator gene and in the modulator site to which the regulator binds ($\tau + \rho$). Mutations in the structural gene for the effector protein could be included by adding the appropriate target size to the target size of the promoter (π), in the case of the negative mode, or the modulator/regulator ($\tau + \rho$), in the case of the positive mode. Similarly, in this study we have emphasized the predominant types of mutations that disrupt normal function. Those that might augment normal function can be considered by again adding their target size to the target size of another mutation that results in a similar phenotype. For example, a mutation in an operator site might result in tighter binding of the cognate repressor and failure to allow induction of gene expression in the high-demand environment. Such a mutant would exhibit the same phenotype as the promoter mutants we have considered. The target size for mutations that augment binding, which is presumably smaller than the target size for mutations that disrupt the normal operator, can be added to the target size for mutations in the promoter (π).

Mutants that result in phenotypes different from those considered here also can be added in a straightforward manner. In these cases, one first calculates the individual threshold for each class of mutation; this may involve entirely different sets of parameters and not just a different target size for mutation. Then one adds these

thresholds to obtain the region of allowable selection for the wild-type regulatory system. For the cases described in the previous paragraph, this method and the method of simply adding the appropriate target sizes produce the same results (data not shown).

In summary, the quantitative development of demand theory reveals unexpected relationships between the demand for gene expression D and the average ON/OFF cycle time for the gene C, which is a manifestation of the organism's life cycle. The theory provides equations for the rate and extent of selection, and these reveal well-defined regions of the C *vs.* D plot within which selection is realizable. The realizable regions for the positive and negative mode exhibit an inherent asymmetry with characteristic values for D_{min}, D_{max}, and C_{max}. The demand theory of gene regulation can be extended within the framework presented here to include organisms with life cycles that are more complex than the two phases illustrated in this article and regulatory systems that are more complex than a single mechanism of gene control.

I thank Drs. S. Cooper, R. G. Freter, D. E. Kirschner, J. V. Neel, and M. S. Swanson for critically reading the manuscript and two anonymous reviewers who made valuable suggestions for clarifying key concepts in the theory. This work was supported in part by U.S. Public Health Service grant RO1-GM30054 from the National Institutes of Health and U.S. Department of Defense grant N00014-97-1-0364 from the Office of Naval Research.

LITERATURE CITED

Datta, A., and S. Jinks-Robertson, 1995 Association of increased spontaneous mutation rates with high levels of transcription in yeast. Science **268:** 1616–1619.

Francino, M. P., L. Chao and M. A. Riley, 1996 Asymmetries generated by transcription-coupled repair in enterobacterial genes. Science **272:** 107–109.

Leclerc, J. E., B. Li, W. L. Payne and T. A. Cebula, 1996 High mutation frequencies among *Escherichia coli* and *Salmonella* pathogens. Science **274:** 1208–1211.

Miller, J. H., and W. S. Reznikoff, 1980 *The Operon.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Press, W. H., B. P. Flannery, S. A. Teukolsky and W. T. Vetterling, 1988 *Numerical Recipes in C.* Cambridge University Press, New York.

Salyers, A. A., 1994 *Bacterial Pathogenesis: A Molecular Approach.* ASM Press, Washington, DC.

Savageau, M. A., 1974 Genetic regulatory mechanisms and the ecological niche of *Escherichia coli.* Proc. Natl. Acad. Sci. USA **71:** 2453–2455.

Savageau, M. A., 1977 Design of molecular control mechanisms and the demand for gene expression. Proc. Natl. Acad. Sci. USA **74:** 5647–5651.

Savageau, M. A., 1979 Autogenous and classical regulation of gene expression: a general theory and experimental evidence, pp. 57–108 in *Biological Regulation and Development,* Vol. 1, edited by R. F. Goldberger. Plenum, New York.

Savageau, M. A., 1983a Regulation of differentiated cell-specific functions. Proc. Natl. Acad. Sci. USA **80:** 1411–1415.

Savageau, M. A., 1983b Models of gene function: general methods of kinetic analysis and specific ecological correlates, pp. 3–25 in *Foundations of Biochemical Engineering: Kinetics and Thermodynamics in Biological Systems,* edited by H. W. Blanch, E. T. Papoutsakis and G. N. Stephanopoulos. American Chemical Society, Washington, DC.

Savageau, M. A., 1985 Coupled circuits of gene regulation, pp. 633–642 in *Sequence Specificity in Transcription and Translation,* edited by R. Calendar and L. Gold. Alan R. Liss, New York.

Savageau, M. A., 1989 Are there rules governing patterns of gene regulation?, pp. 42–66 in *Theoretical Biology—Epigenetic and Evolutionary Order,* edited by B. C. Goodwin and P. T. Saunders. Edinburgh University Press, Edinburgh.

Savageau, M. A., 1998 Demand theory of gene regulation. II. Quantitative application to the lactose and maltose operons of *Escherichia coli.* Genetics **149:** 1677–1691.

Schwartz, M., 1987 The maltose regulon, pp. 1482–1502 in *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology,* edited by F. C. Neidhardt. American Society for Microbiology, Washington, DC.

Slack, J. M. W., 1992 *From Egg to Embryo,* Ed. 2. Cambridge University Press, Cambridge.

Communicating editor: R. H. Davis

Copyright © 1998 by the Genetics Society of America

Demand Theory of Gene Regulation. II. Quantitative Application to the Lactose and Maltose Operons of *Escherichia coli*

Michael A. Savageau

Department of Microbiology and Immunology, The University of Michigan, Ann Arbor, Michigan 48109-0620

Manuscript received December 15, 1997
Accepted for publication May 6, 1998

ABSTRACT

Induction of gene expression can be accomplished either by removing a restraining element (negative mode of control) or by providing a stimulatory element (positive mode of control). According to the demand theory of gene regulation, which was first presented in qualitative form in the 1970s, the negative mode will be selected for the control of a gene whose function is in low demand in the organism's natural environment, whereas the positive mode will be selected for the control of a gene whose function is in high demand. This theory has now been further developed in a quantitative form that reveals the importance of two key parameters: cycle time *C*, which is the average time for a gene to complete an ON/OFF cycle, and demand *D*, which is the fraction of the cycle time that the gene is ON. Here we estimate nominal values for the relevant mutation rates and growth rates and apply the quantitative demand theory to the lactose and maltose operons of *Escherichia coli*. The results define regions of the *C vs. D* plot within which selection for the wild-type regulatory mechanisms is realizable, and these in turn provide the first estimates for the minimum and maximum values of demand that are required for selection of the positive and negative modes of gene control found in these systems. The ratio of mutation rate to selection coefficient is the most relevant determinant of the realizable region for selection, and the most influential parameter is the selection coefficient that reflects the reduction in growth rate when there is superfluous expression of a gene. The quantitative theory predicts the rate and extent of selection for each mode of control. It also predicts three critical values for the cycle time. The predicted maximum value for the cycle time *C* is consistent with the lifetime of the host. The predicted minimum value for *C* is consistent with the time for transit through the intestinal tract without colonization. Finally, the theory predicts an optimum value of *C* that is in agreement with the observed frequency for *E. coli* colonizing the human intestinal tract.

THE life cycle of a microbe, in the simplest case, consists of alternative phases. The demand for expression of some effector genes will be high in one phase and low in the other, and adapting the level of expression to this varying demand requires a functional regulatory mechanism. It has long been known that the same regulatory function, for example, induction of gene expression, can be accomplished in one of two different modes: the negative mode involves the removal of a restraining element, which permits expression from a high-level promoter, whereas the positive mode involves the provision of a stimulatory element, which facilitates expression from a low-level promoter. The demand theory of gene regulation provides a selectionist explanation for this fundamental duality (*e.g.*, see SAVAGEAU 1977, 1989).

The components of a minimal regulatory mechanism consist of a promoter site, a modulator site, and a regulator gene encoding the protein that binds the modulator site in response to environmental cues. Each of these components is subject to a rate of mutation that is determined by the number of critical bases in its nucleotide sequence and the mutation rate per base per round of DNA replication. A mutant altered in one of these components may exhibit two different phenotypes depending upon the phase of the life cycle in which it is expressed. The growth rate of the organism serves as the relevant phenotype, and selection is based upon differences in growth rate among wild-type and mutant organisms.

The quantitative development of demand theory (SAVAGEAU 1998) combines these elements of life cycle, ecology, physiology, and molecular genetics to predict regions of the *C vs. D* plot within which selection for the wild-type regulatory mechanisms is realizable. These regions define minimum and maximum values for demand. This theory ties together a number of important variables, including growth rates, mutation rates, and minimum and maximum demands for gene expression. We apply demand theory here to the lactose (*lac*) and maltose (*mal*) catabolic systems of *Escherichia coli*, and we show that this theory also yields predictions for the rate and extent of selection and for minimum, maximum, and optimal cycle times of *E. coli* that are in reasonable agreement with independent experimental data.

Life cycle of *E. coli*: The normal life cycle of an organ-

Address for correspondence: 5641 Medical Science Bldg. II, Department of Microbiology and Immunology, The University of Michigan Medical School, Ann Arbor, MI 48109-0620.
E-mail: savageau@umich.edu

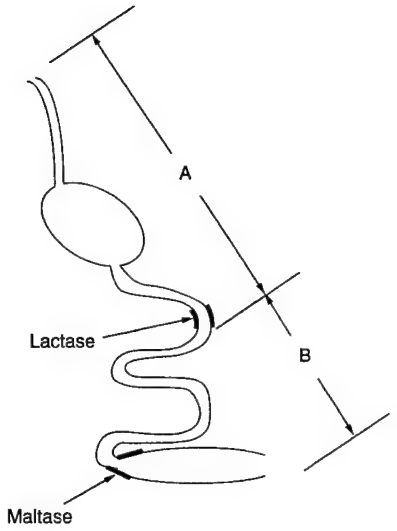


FIGURE 1.—The life cycle of *Escherichia coli* alternates between two different environments, the proximal portions of the digestive tract (A), where lactose levels are relatively high and maltose levels relatively low, and the distal portions (B), where lactose levels are relatively low and maltose levels relatively high. In the environment labeled A, there is a high demand for expression of the A-specific *lac* genes and a low demand for expression of the B-specific *mal* genes, whereas in the alternative environment labeled B the demand for these same genes is reversed. As described in the text, the average time required for the organism to complete its life cycle is denoted by C, which also represents the average time for the A-specific (or B-specific) genes to complete their ON/OFF cycle. The fraction of this cycle time spent in environment A (or B), which also represents the demand for expression of the A-specific (or B-specific) genes, is denoted by D.

ism defines the demand for expression of its effector genes. The life cycle of *E. coli* as it passes from one host to another will be considered here in terms of two different environments (Figure 1). The first will be identified with the proximal portions of the digestive tract for a lactose-tolerant host that ingests both lactose and starch (which consists largely of maltose). This is the environment in which rapid growth occurs during the transition between stable association with one host and then another. The second environment will be identified with the distal end of the small intestine and the colon of the host in which colonization and slow growth take place. Also included in the second environment will be the host's surroundings through which the bacteria pass to enter a subsequent host. This is admittedly a simplification of a more complex ecology (COOKE 1974; FRETER 1976; SAVAGEAU 1983), but nevertheless it captures the essential features for our purposes here. Additional environments and more complex linkages among them in principle can be handled by the same methods.

Ecology and gene expression: We shall consider the *lac* operon as representative of a low-demand function governed by the negative mode of control (MILLER and REZNIKOFF 1980) and the *mal* operon as representative of a high-demand function governed by the positive mode of control (SCHWARTZ 1987). These systems are well studied at the molecular level, and the evidence regarding their mode of control is clear.

Evidence regarding demand for expression of the *lac* and *mal* operons of *E. coli* comes from studies of intestinal ecology. Lactose is a relatively rare sugar in nature (SHALLENBERGER 1974). The host's lactase enzymes, which hydrolyze this disaccharide and thereby permit its utilization by the host, are located in the proximal small intestine and are subject to developmental regulation (DAHLQVIST 1961; KOLDOVSKY and CHYTIL 1965; WALKER 1968). In contrast, maltose, the breakdown product of all dietary starch, is among the most abundant sugars (WIDDAS 1971). The host's maltase enzymes, which hydrolyze this disaccharide and thereby permit its utilization by the host, are located at the distal end of the small intestine and in the colon (DAHLQVIST 1961; ROSENSWEIG and HERMAN 1968). This information suggests that the *lac* operon of *E. coli* is likely to be expressed at high levels in the first environment and at low levels in the second, whereas the *mal* operon is likely to be expressed at high levels in the second environment and at low levels in the first.

The time required for *E. coli* to pass through the high-demand environment for lactose utilization is about 3 hr. This is one-half the average time required to reach the colon (MADSEN 1992); the 3-hr figure is also based on measured patterns of lactose utilization (BOND and LEVITT 1976; MALAGELADA *et al.* 1984). Much of the ingested lactose is hydrolyzed to constituent sugars that are absorbed by the host in the proximal small intestine; the remainder is catabolized by the bacteria so that very little lactose normally reaches the colon (BOND and LEVITT 1976). From this 3-hr figure for time in the high-demand environment, one can predict that the cycle time of *E. coli* will be inversely related to the demand for expression of its lactose operon $C = 3/D$.

We estimate the time for passage through the low-demand environment for maltose utilization to be ~6 hr. This is the average time required for a bolus of ingested food to reach the distal portions of the small intestine and colon (MADSEN 1992). We assume that free maltose is sparse in the proximal portions of the small intestine and that it becomes abundant only in the distal portion of the intestinal tract where the host's maltase enzymes are localized (DAHLQVIST 1961; ROSENSWEIG and HERMAN 1968). From this 6-hr figure for time in the low-demand environment, one can predict that the cycle time of *E. coli* will be inversely related to 1 minus the demand for expression of its maltose operon $C = 6/(1 - D)$.

ESTIMATION OF PARAMETER VALUES

The demand theory of gene regulation involves three levels of parameters (SAVAGEAU 1998): constituent pa-

TABLE 1

Definitions and nominal values for the constituent parameters that determine the growth rates and mutation rates for organisms with positive and negative modes of control in high- and low-demand environments

Symbol	Definition	Nominal value[a]	
		Negative control	Positive control
μ	Reference mutation rate	6E-10 base^{-1} generation^{-1}	6E-10 base^{-1} generation^{-1}
π	Mutation rate, relative to μ, for loss of a promotor with negative control	10	—
υ	Mutation rate, relative to μ, for gain of an up-promoter with positive control	—	1
τ	Mutation rate, relative to μ, for loss of a regulator's functional target site	20	20
ρ	Mutation rate, relative to μ, for loss of a functional regulator protein	60	60
ε	Mutation rate, relative to μ, when expression is increased 100-fold	1	1
γ	Reference growth rate in the nutritionally richer of the two environments	1.0 generation hour^{-1}	1.0 generation hour^{-1}
δ	Growth rate, relative to γ, when in the more nutritionally deficient environment	0.0125	0.0125
λ	Growth rate, relative to γ, when there is a loss of expression with negative control	0.97	—
λ	Growth rate, relative to $\delta\gamma$, when there is loss of expression with positive control	—	0.97
σ	Growth rate, relative to $\delta\gamma$, when there is superfluous expression with negative control	0.999	—
σ	Growth rate, relative to γ, when there is superfluous expression with positive control	—	0.999

[a] See text for estimation of parameter values.

rameters, individual growth rates and mutation rates, and macroscopic rate constants. Estimates for the values of the constituent parameters in our model are given below. In a subsequent section we will determine the consequences of these choices by examining other values for each of the parameters. As we shall see, the exact values are not critical for about half of the parameters, whereas the values for two of them are extremely influential.

Reference mutation rate, μ: For *E. coli*, the spontaneous mutation rate is estimated to have a nominal value of μ_0 = 6E–10 per base per DNA replication (DRAKE 1991). The spontaneous mutation rate for loss of function in the modulator or promoter sites of our model can be determined from estimates of the spontaneous mutation rate per base and the number of critical bases that define these sites.

Relative mutation rate for loss of a high-level promoter site, π: Promoters encompass a region of ~75 nucleotides upstream of the RNA start site (HARLEY and REYNOLDS 1987; LISSER and MARGALIT 1993). Although most of the information in promoter sequences is localized within two 6-base blocks with variable spacing between them (the "consensus" elements at positions −10 and −35 relative to the start site), there is only limited base conservation at most positions. If we assume that each of 10 nucleotides is critical for the definition of a high-level promoter (SCHNEIDER *et al.* 1986), then the relative mutation rate is π_0 = 10.

Relative mutation rate for gain of a high-level promoter site, υ: Spontaneous up-promoter mutations with the positive mode occur at about one-tenth the frequency of spontaneous down-promoter mutations with the negative mode (G. GUSSIN, personal communication). If we assume that a low-level promoter can be converted to a high-level promoter by a single mutation in a critical base, then the relative mutation rate for this gain of a high-level promoter site is υ_0 = 1.

Relative mutation rate for loss of a regulator's functional target site, τ: Targets for the binding of regulator proteins are the modulator sites—operator sites in the case of the negative mode and initiator sites in the case of the positive mode. Operator sites span a region of ~100 nucleotides upstream of the RNA start site (REALLA and COLLADO-VIDES 1996). If we assume that each of 20 nucleotides is critical for the definition of the operator (SCHNEIDER *et al.* 1986), then the relative mutation rate in this case is τ_0 = 20. Although the sizes of initiator sites are about one-half those of operator sites, 50 nucleotides upstream of the RNA start site

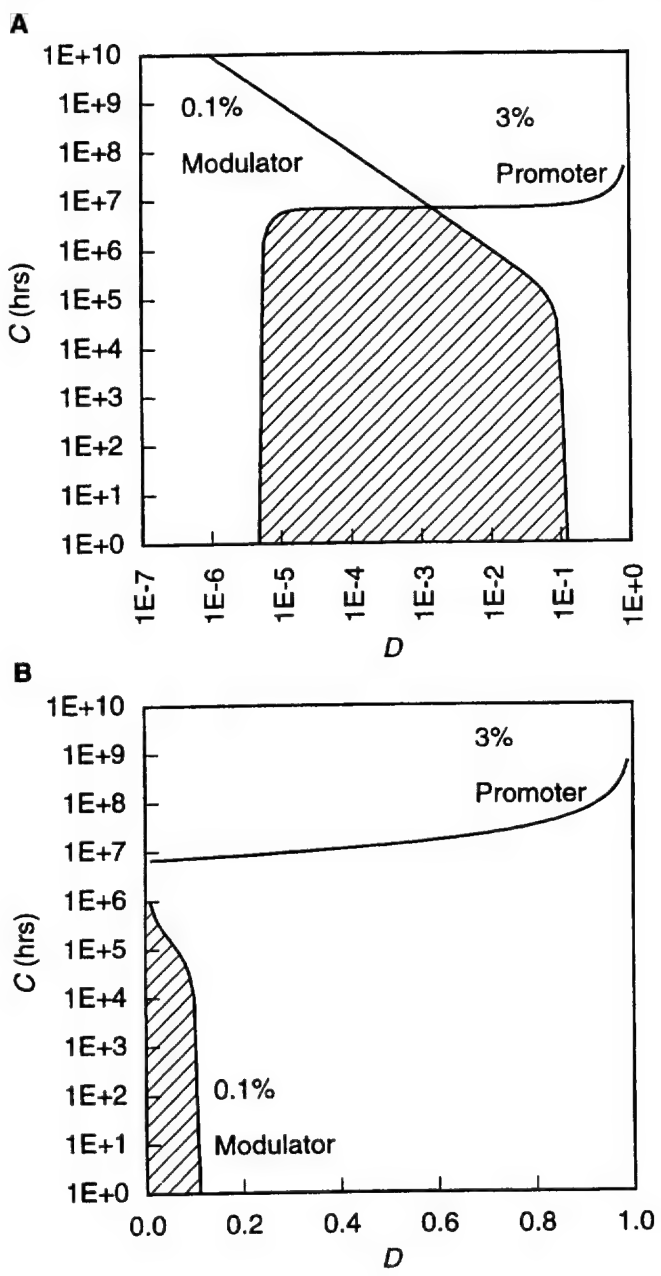


FIGURE 2.—Thresholds for selection of a wild-type regulatory mechanism with a negative mode. (A) The demand is represented with a logarithmic scale, and one sees that D_{min} = 4.8E–6 and D_{max} = 0.1. (B) The demand is represented with a linear scale. Results are shown for the nominal values of the parameters in Table 1. See text for discussion.

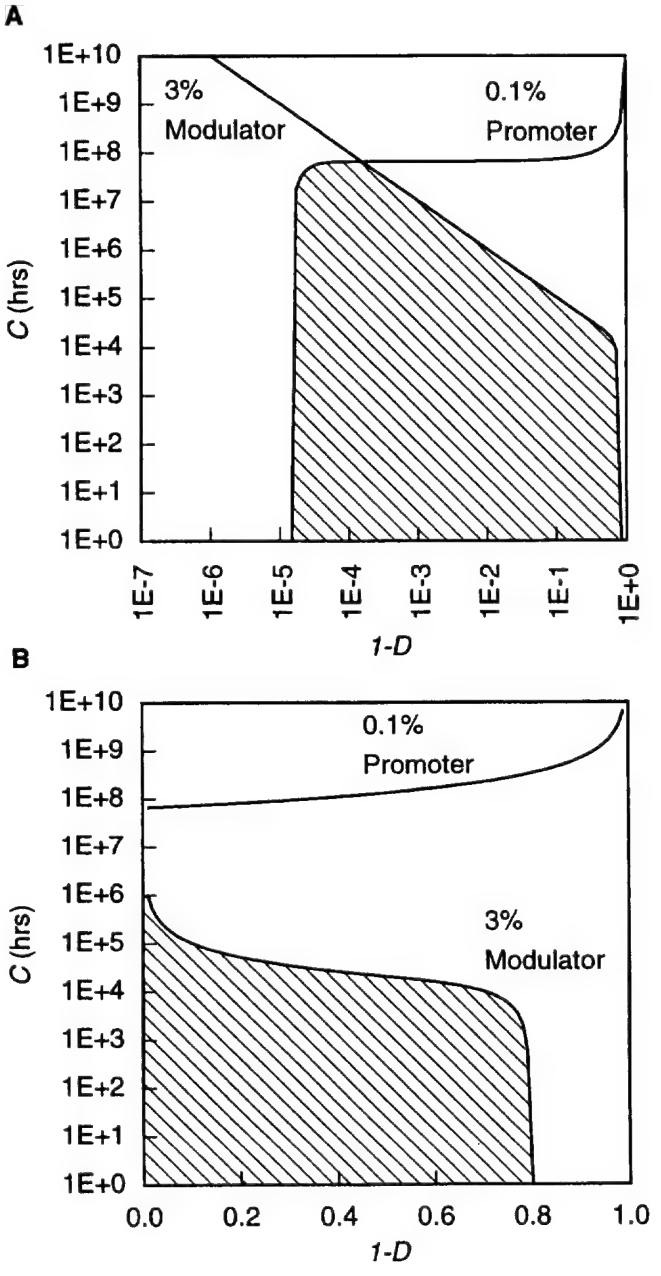


FIGURE 3.—Thresholds for selection of a wild-type regulatory mechanism with a positive mode. (A) The demand is represented with a logarithmic scale, and one sees that D_{min} = 1–0.8 and D_{max} = 1–1.5E–5. (B) The demand is represented with a linear scale. Results are shown for the nominal values of the parameters in Table 1. See text for discussion.

(REALLA and COLLADO-VIDES 1996), the information needed to locate these sites within the genome is similar to that for operators (SCHNEIDER *et al.* 1986). If we assume that each of 20 nucleotides also is critical for the definition of the initiator, then the relative mutation rate in this case is τ_0 = 20.

Relative mutation rate for loss of a functional regulator protein, ρ: We shall assume that a typical regulator protein has 30 amino acid residues that are critical for binding to its modulator site (operator in the case of the negative mode or initiator in the case of the positive mode) and for properly affecting transcription initiation. This implies that the regulator gene has ~60 bases that are critical because the identity of the base in the third codon position is largely irrelevant. Thus, we obtain a relative mutation rate of ρ_0 = 60 for loss of a functional regulator protein.

Relative mutation rate as a function of gene expres-

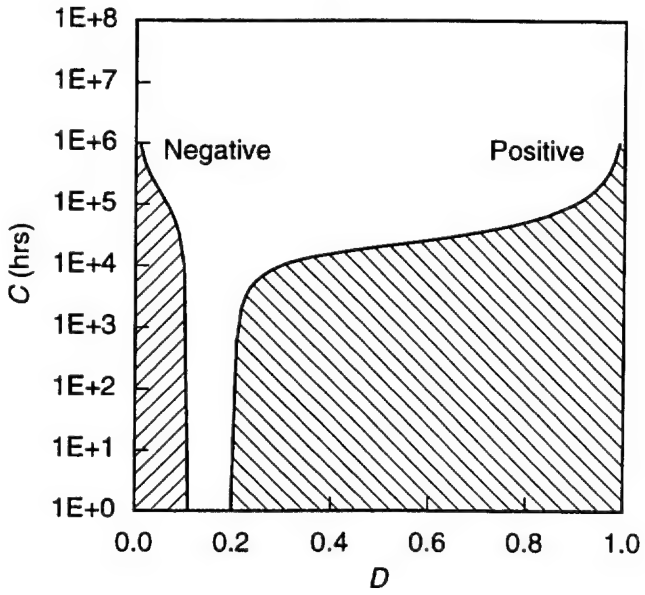


FIGURE 4.—Thresholds for discriminate selection of wild-type regulatory mechanisms with negative or positive modes. The results in Figures 2B and 3B are shown here on the same axes, where it is clear that the D_{max} for selection of the negative mode is less than the D_{min} for selection of the positive mode.

sion, ε: There is evidence to suggest that the rate of spontaneous mutation increases with the rate at which the DNA is being transcribed (DATTA and JINKS-ROBERTSON 1995). There also is evidence to suggest that the rate of spontaneous mutation decreases because of transcription-coupled repair mechanisms (FRANCINO *et al.* 1996). This is one of the questions we wish to examine further in our quantitative analysis. We initially assume that there is no significant effect on the mutation rate either way; therefore, we assign a value of $\varepsilon_0 = 1$ for this relative mutation rate. If one were to assume a change in mutation rate that is proportional (or inversely proportional) to the rate of transcription, then the mutation rate relative to the reference would be given by $\varepsilon = k \times 100$ (or $k/100$), where k is the proportionality constant. (Recall that the capacity for regulation is assumed to be 100 and that expression is assumed to be fully ON or fully OFF.)

Reference growth rate, γ: We shall assume that *E. coli* grows with a doubling time of 1 hr in the nutritionally richer of the two environments; thus, the nominal value for the reference growth rate is $\gamma_0 = 1.0$. This is not an unreasonable value because it is known that bacteria like *E. coli* can double in a period as short as 20 min (MAALØE and KJELDGAARD 1965). In any case, the simple value of unity provides a convenient reference; should the actual value be different, one can simply rescale the time accordingly, and none of our results would change.

Relative growth rate with loss of normal expression, λ: Because expression is either fully ON or fully OFF and the capacity for regulation is 100, which supports the nominal growth rate, a failure of expression is assumed to result in a basal level of expression, which would support only a 100-fold reduction in growth rate if there were no other carbon source in the environment. However, in the complex environment of the intestinal tract there are multiple carbon sources, and the reduction in growth rate will therefore be less. We shall assume a 3% reduction in growth rate. Thus, the nominal value for this parameter is set at $\lambda_0 = 0.97$.

Relative growth rate with superfluous expression, σ: When the demand is such that a function is normally turned OFF and a regulatory mutation causes the function to be fully expressed under inappropriate circumstances, the cell unnecessarily expends resources for material and energy. Experimental evidence in the case of β-galactosidase expression in *E. coli* (NOVICK and WEINER 1957; KOCH 1983) suggests that such inappropriate expression decreases the growth rate by <1%; we shall assume a 0.1% reduction. The growth rate, relative to the reference growth rate, is thus assigned a nominal value of $\sigma_0 = 0.999$.

Relative growth rate in the more nutritionally deficient of the two environments, δ: From measurements of the mean transit time through the human intestinal tract (CUMMINGS and WIGGINS 1976; GEAR *et al.* 1980), and the assumption that it is a well-stirred chemostat, one can calculate that the average doubling time for net growth of *E. coli* in the intestinal tract is about 40 hr (SAVAGEAU 1983). Because the intestinal tract is not a well-stirred chemostat, but rather a very heterogeneous environment in which the growth is undoubtedly faster in the proximal regions and slower in the distal, the doubling time of *E. coli* in the more deficient distal environment will be longer than the average. There are no good measurements to go by, so we will arbitrarily set the doubling time for growth in the more deficient environment to be two times the average value given above. Thus, the growth rate in the more deficient environment, relative to the reference growth rate in the richer environment, is given by a nominal value of $\delta_0 = 0.0125$.

Criterion for selection, θ: Our criterion for selection is that each mutant population shall be reduced to no more than 0.05% of the wild-type population or, alternatively, that the sum of the two mutant populations shall be reduced to no more than 0.1% of the wild-type population. This is similar to values that are found in the literature (LECLERC *et al.* 1996). Thus, the criterion for selection is assigned a nominal value of $\theta_0 = 0.0005$.

Estimation of macroscopic parameters: The values of the macroscopic parameters in each environment and for each mode of control are determined as follows. First, the constituent parameters given above are combined to represent the relevant growth rates (g_w, g_p, g_m, g_a). The growth rate of the wild-type organism g_w in the first environment is γ (the reference), and in the second

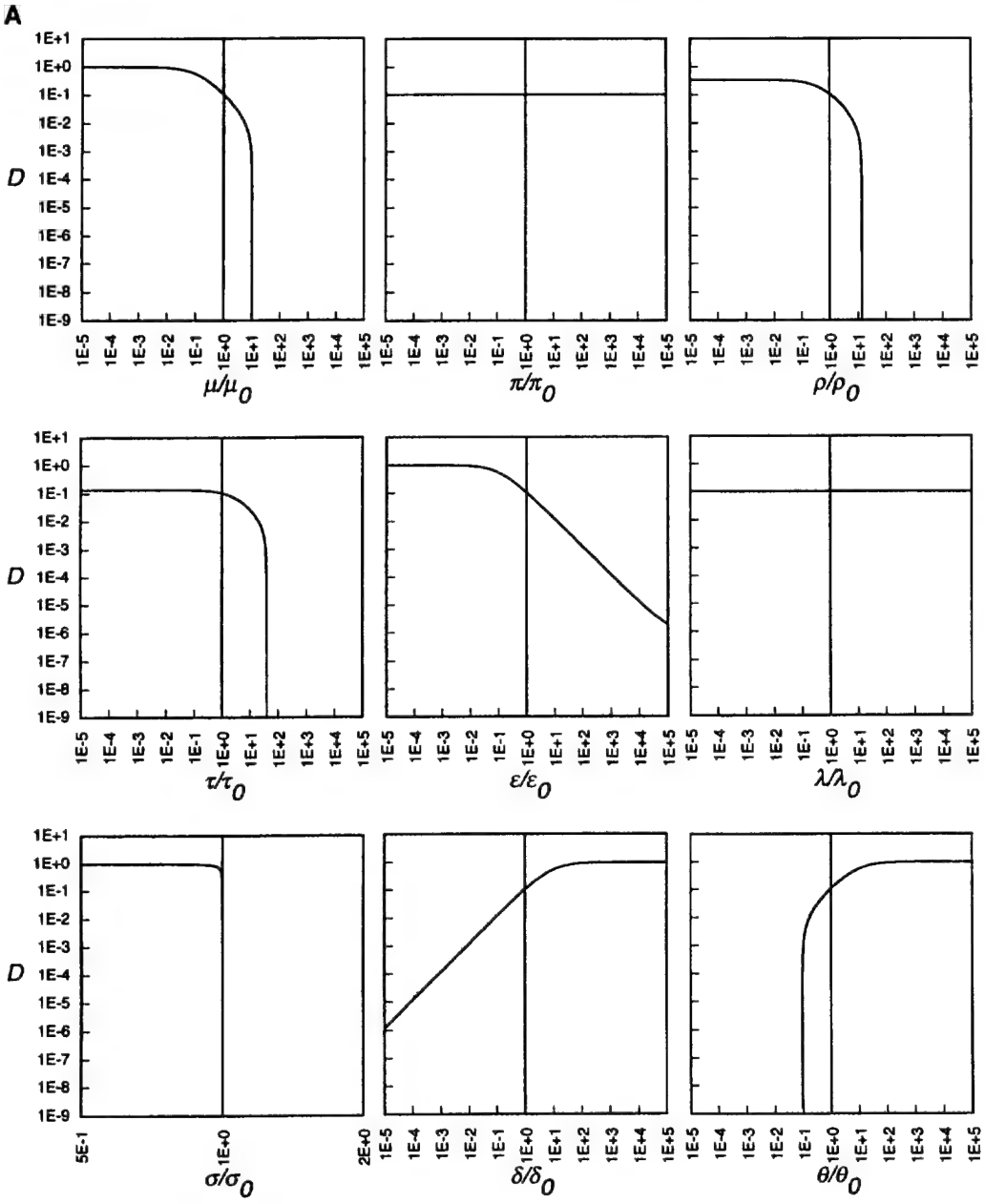


FIGURE 5.—Influence of the constituent parameters on the values for D_{min} and D_{max}. Each parameter is varied about its nominal value, and the resulting values for D_{min} and D_{max} are calculated. Results for the parameter γ are not shown because it exhibits no influence in all cases. (A) D_{max} for the negative mode of control. (B) D_{min} for the negative mode of control. (C) D_{min} for the positive mode of control. (D) D_{max} for the positive mode of control. The local parameter sensitivities are summarized in Table 2.

it is γ multiplied by δ, the relative growth rate in the nutritionally deficient environment. The growth rates of the promoter mutants g_p, modulator mutants g_m, and promoter/modulator (double) mutants g_d in the two environments are the same as those of the wild type, but multiplied when appropriate by relative growth rates that reflect either loss of expression that is normally ON (λ) or superfluous expression that is normally OFF (σ). For example, the growth rate of a *lac* modulator mutant (g_m) is γ in the first environment, where its pattern of gene expression mimics that of the wild type, and $\gamma\delta\sigma$ in the second, where expression of the *lac* operon is superfluous. The growth rate of a *mal* modulator mutant is γ in the first environment, where its pattern of gene expression mimics the wild type, and $\gamma\delta\lambda$ in the second, where there is a failure to express the *mal* operon.

Second, the constituent parameters are combined to represent the mutation rates between populations (m_{pw}, m_{mw}, m_{dp}, m_{dm}). Each mutation rate is given by the product of the number of critical bases that define the structure in question (π, υ, τ, or ρ), the spontaneous mutation rate per base per DNA replication (μ), and a factor reflecting transcription-related mutation or repair (ε) when appropriate. For example, the rate of production of *lac* promoter mutants from wild-type organisms (m_{pw}) is $\pi\mu\varepsilon$ in the first environment, where the *lac* operon is being actively transcribed, and $\pi\mu$ in the second, where it is not. The rate of production of *mal* promoter mutants from wild-type organisms is $\upsilon\mu$

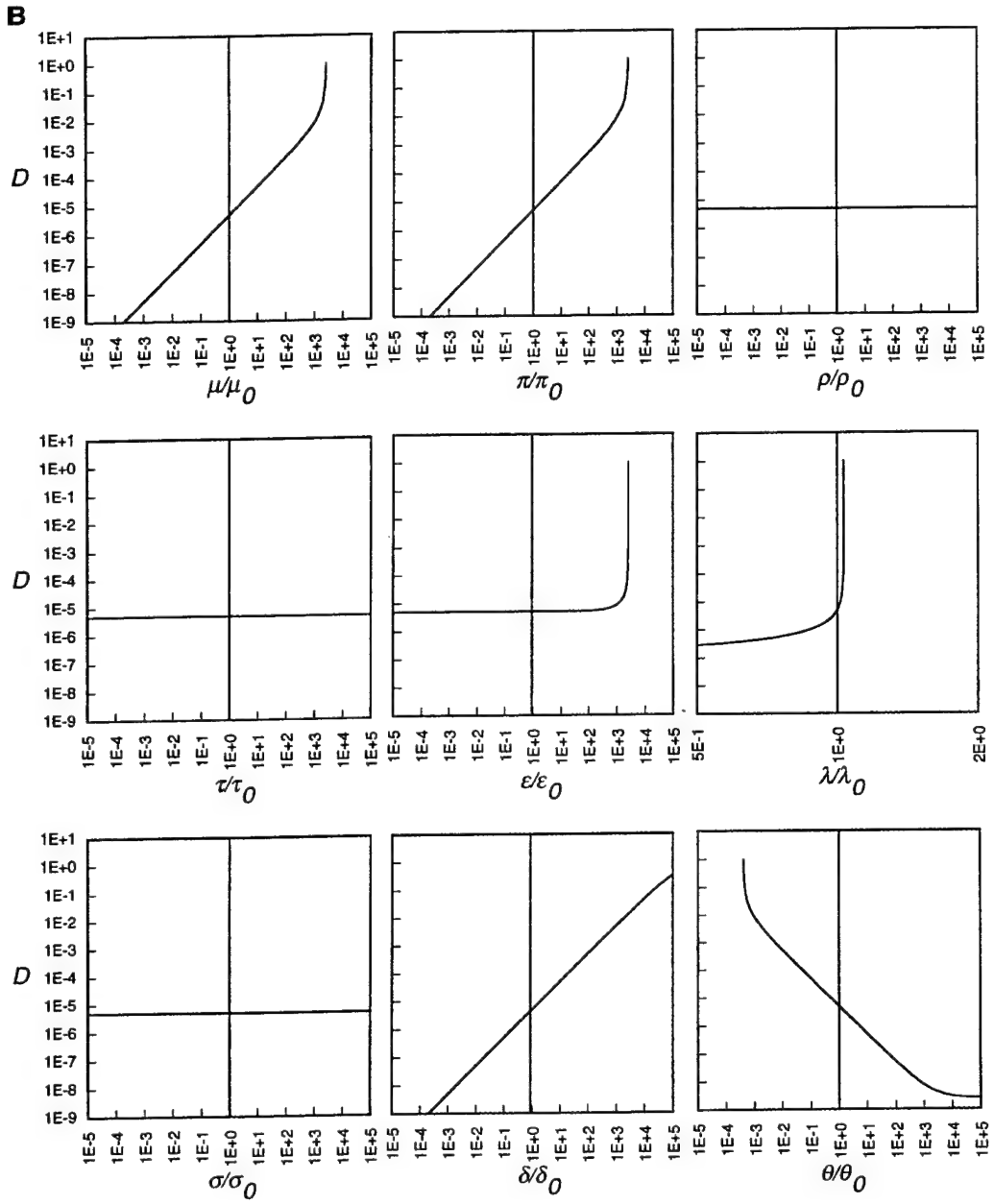


FIGURE 5.—*Continued.*

in the first environment, where the *mal* operon is not being transcribed, and υμε in the second, where it is.

Finally, the growth rates and mutation rates are combined to represent the macroscopic rate-constant parameters that characterize the population dynamics (α_{ww}, α_{pp}, α_{pw}, α_{mm}, α_{mw}, α_{dd}, α_{dp}, α_{dm}). For example, the rate constant for net growth of the promoter mutant α_{pp} is given by its intrinsic growth rate g_p minus the rate of loss due to the production of double mutants, which is given by the mutation rate per DNA replication m_{dp} times the intrinsic growth rate of the promoter mutant g_p. The rate constant for production of promoter mutants from the wild-type population α_{pw} is given by the mutation rate per DNA replication m_{pw} times the intrinsic growth rate of the wild-type organism g_w. The other rate-constant parameters are determined in a similar fashion. Thus, $\alpha_{ww} = [1 - (m_{pw} + m_{mw})]g_w$, $\alpha_{pp} = (1 - m_{dp})g_p$, $\alpha_{pw} = m_{pw}g_w$, $\alpha_{mm} = (1 - m_{dm})g_m$, $\alpha_{mw} = m_{mw}g_w$, $\alpha_{dd} = g_d$, $\alpha_{dp} = m_{dp}g_p$, $\alpha_{dm} = m_{dm}g_m$.

SELECTION OF WILD-TYPE REGULATORY MECHANISMS

Determination of the thresholds for selection: The threshold for selection of a wild-type promoter or modulator is determined as follows. As described above, the constituent parameters are combined to represent the relevant growth rates and mutation rates that enter into the macroscopic parameters that characterize the population dynamics of mutant and wild-type organisms. The population dynamic equations are solved in the two environments to yield an equation for the ratio of mu-

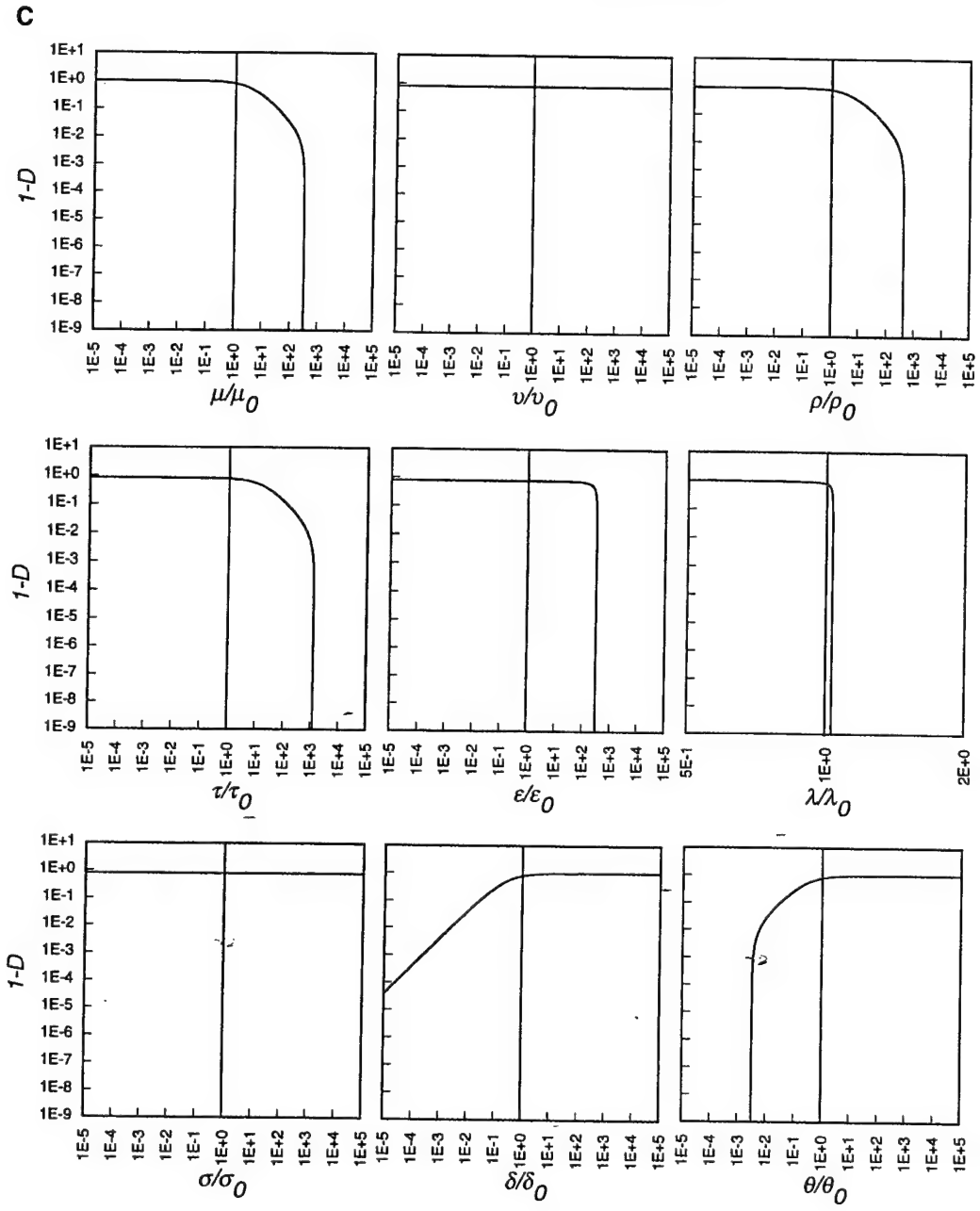


FIGURE 5.—*Continued.*

tant to wild-type population numbers. This ratio is set equal to θ, the criterion for selection, and the resulting equation expresses the relationship between cycle time C and the demand for gene expression D that constitutes the threshold for selection.

The threshold for selection of a wild-type promoter or modulator (regulator) can be obtained by solving the threshold equation for C as a function D using the method of bisection (PRESS *et al.* 1988). These numerically calculated thresholds and the analytically determined asymptotes (SAVAGEAU 1998) are nearly indistinguishable, except in the region of transition between low- and high-C asymptotes. Only those values of C and D that lie in the region of overlap below both the promoter and modulator thresholds will allow selection of the wild-type regulatory mechanism.

Regions in which selection for negative and positive modes is realizable: When nominal values are assumed for the parameters of the model (see Table 1), one finds the thresholds plotted in Figures 2 and 3. The thresholds shown in Figure 2 for the negative mode of control (3% selection coefficient against the promoter mutant and 0.1% against the modulator mutant) exhibit a narrow region of overlap. In contrast, the thresholds shown in Figure 3 for the positive mode of control (0.1% selection coefficient against the promoter mutant and 3% against the modulator mutant) exhibit a wide region of overlap. As predicted (SAVAGEAU 1998), the positive and negative modes are associated with asymmetric regions of the C *vs.* D plot in which selection is realizable. Note that the horizontal axis in the case of the positive mode is plotted as values of $1 - D$, instead of D; this allows

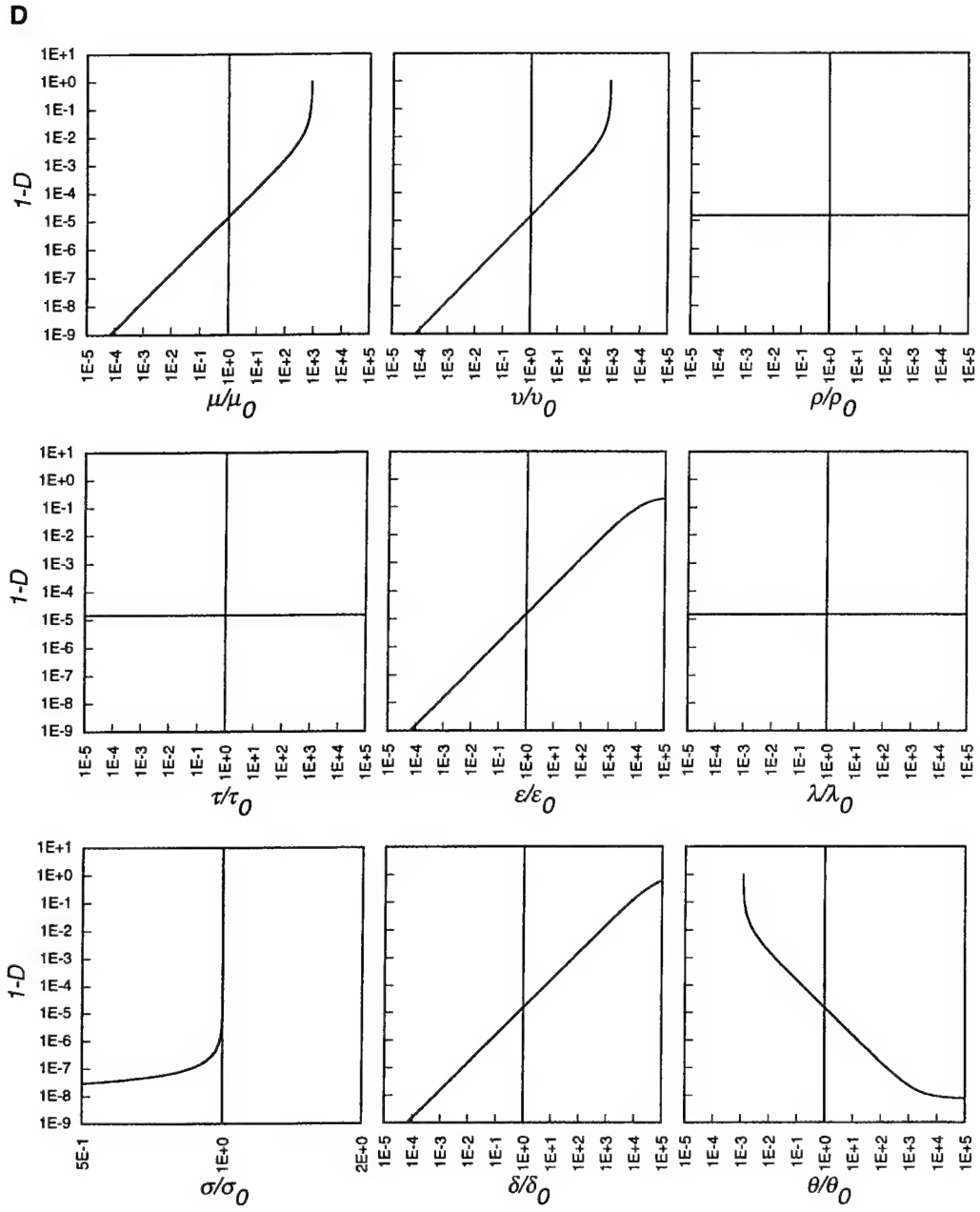


FIGURE 5.—*Continued.*

us to distinguish more clearly the threshold values near unity when plotting demand in logarithmic coordinates. The extent of the asymmetry is perhaps more evident when the thresholds for selection against the modulator mutants are plotted on the same scale for both the negative and positive modes (Figure 4). The regions for which selection is realizable are nonoverlapping, which indicates discriminate selection for the positive and negative modes.

Influence of parameters on minimum and maximum values for demand: Selection requires the demand for gene expression to be greater than the asymptote for the minimum threshold and less than the asymptote for the maximum threshold, that is, $D_{min} < D < D_{max}$. The parameters in our model influence these asymptotic values to various degrees. It is important to examine a range of values for each of the parameters because there is some uncertainty in the nominal values for many of them. We have systematically varied each parameter about its nominal value given in Table 1 and observed the resulting changes in the minimum and maximum values for demand. The results are shown in Figure 5.

Five classes of influence can be discerned in Figures 5A–5D. First, in many cases there is no discernible influence [Figure 5A (π and λ), Figure 5B (ρ, τ, ε, and σ), Figure 5C (υ, τ, ε, and σ), and Figure 5D (ρ, τ, and λ)]. Second, in several cases there is a nearly linear variation with the change in parameter value [Figure 5A (μ, ρ, ε, δ, and θ), Figure 5B (μ, π, δ, and θ), and Figure 5D (μ, υ, ε, δ, and θ)]. Third, in five cases there is a nearly cube-root influence [Figure 5A (τ) and Figure 5C (μ, ρ, δ, and θ)]. Fourth, in two cases there is a

TABLE 2

Influence of the constituent parameters on the values of D_{min} and D_{max}

	Parameter sensitivities ($\partial \log \cdot / \partial \log p$)[a]			
	Negative mode		Positive mode	
Parameter, p	D_{min}[b]	D_{max}[c]	$1 - D_{min}$[d]	$1 - D_{max}$[e]
μ	1.00	−0.990	−0.205	1.00
π	1.00	0.000	—	—
υ	—	—	0.000	1.00
τ	0.000	−0.327	−0.0513	0.000
ρ	0.000	−0.769	−0.154	0.000
ε	0.000	−0.895	0.000	1.00
γ	0.000	0.000	0.000	0.000
λ[f]	32.3	0.000	−6.63	0.000
σ[g]	0.000	−989	0.000	1000
δ	1.00	0.895	0.208	1.00
θ	−1.00	0.989	0.205	−1.00

[a] Sensitivities are calculated with D for the negative mode and with $1 - D$ for the positive mode (see text for discussion).

[b] D_{min} is determined by the threshold for selection of the wild-type promoter.

[c] D_{max} is determined by the threshold for selection of the wild-type modulator-repressor interaction.

[d] D_{min} is determined by the threshold for selection of the wild-type modulator-activator interaction.

[e] D_{max} is determined by the threshold for selection of the wild-type promoter.

[f] For the negative mode, λ represents the growth rate of the promoter mutant relative to the wild type; for the positive mode, λ represents the growth rate of the modulator mutant relative to the wild type.

[g] For the negative mode, σ represents the growth rate of the modulator mutant relative to the wild type; for the positive mode, σ represents the growth rate of the promoter mutant relative to the wild type.

moderate (order of magnitude) amplification of the response to a change in parameter value [Figure 5B (λ) and Figure 5C (λ)]. Finally, in two cases there is an extreme (1000-fold) amplification of the response to a change [Figure 5A (σ) and Figure 5D (σ)]. The results obtained for the negative and positive modes exhibit different patterns. The influences in the local region about the nominal values can be summarized numerically by the parameter sensitivities (Shiraishi and Savageau 1992), as shown in Table 2.

From these results one can see that the most influential parameter is the selection coefficient that reflects the diminished growth rate of the organism when there is superfluous gene expression. For the negative mode of gene control, this corresponds to the diminished growth rate of the modulator (regulator) mutants that express the effector function constitutively when it should be OFF. For the positive mode, this corresponds to the diminished growth rate of the promoter mutants that express the effector function at a high level when it should be OFF.

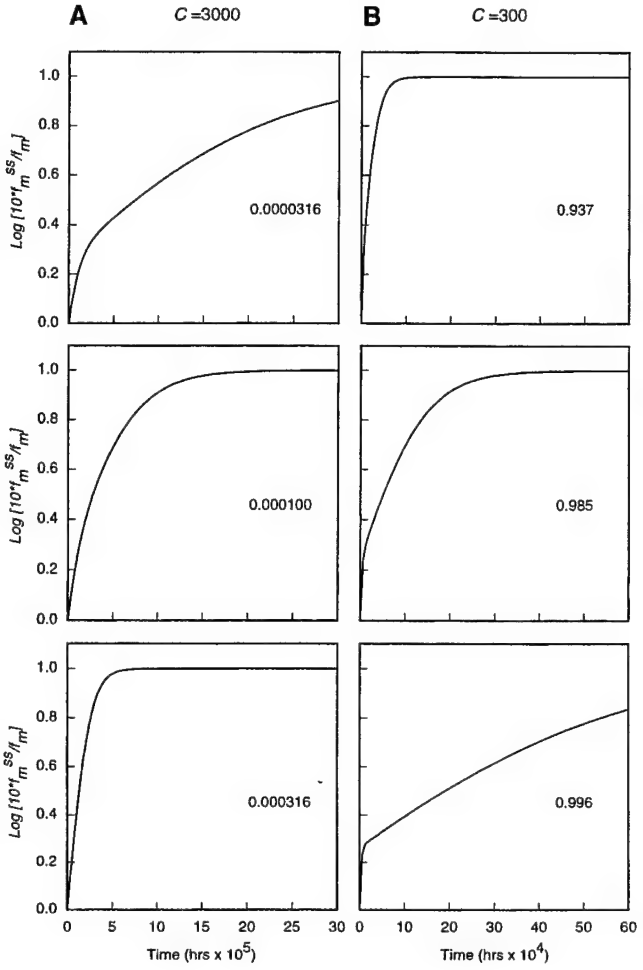


Figure 6.—Enrichment of the wild-type regulatory mechanism with time. The numbers of wild-type, modulator-mutant, and promoter-mutant organisms are represented by the variables X_w,X_m, and X_p. Initially, the ratio $X_w/(X_m + X_p)$, which is the reciprocal of the mutant fraction f_m, is one-tenth of its steady-state (ss) value, and the two types of mutants are equally abundant. The ratio is normalized and plotted as a function of time in units of nC (see text for discussion). The normalized ratio is given by $10 \times [X_w/(X_m + X_p)]/[X_w/(X_m + X_p)]^{ss}$ or $10 f_m^{ss}/f_m$ so that its values vary between 0 and 1 on a logarithmic scale. Time courses are shown for various values of demand D. (A) The negative mode of gene control with a cycle time of $C = 3000$ hr. (B) The positive mode of gene control with a cycle time of $C = 300$ hr.

Time course of selection: The numbers of wild-type, modulator-mutant, and promoter-mutant organisms are represented by the variables X_w, X_m, and X_p. The ratio $X_w/(X_m + X_p)$ is equal to the reciprocal of the *mutant fraction*, which we define as f_m. If we start with equal numbers for the two types of mutants and a ratio $X_w/(X_m + X_p)$ that is one-tenth of its steady-state value, then the enrichment of the wild-type regulatory mechanism with time is obtained from the solution of the population dynamic equations (Equation 9 in Savageau 1998). Given the nominal values for the parameters in Table 1, the time course of selection for various values of demand D is as shown in Figure 6. The temporal behav-

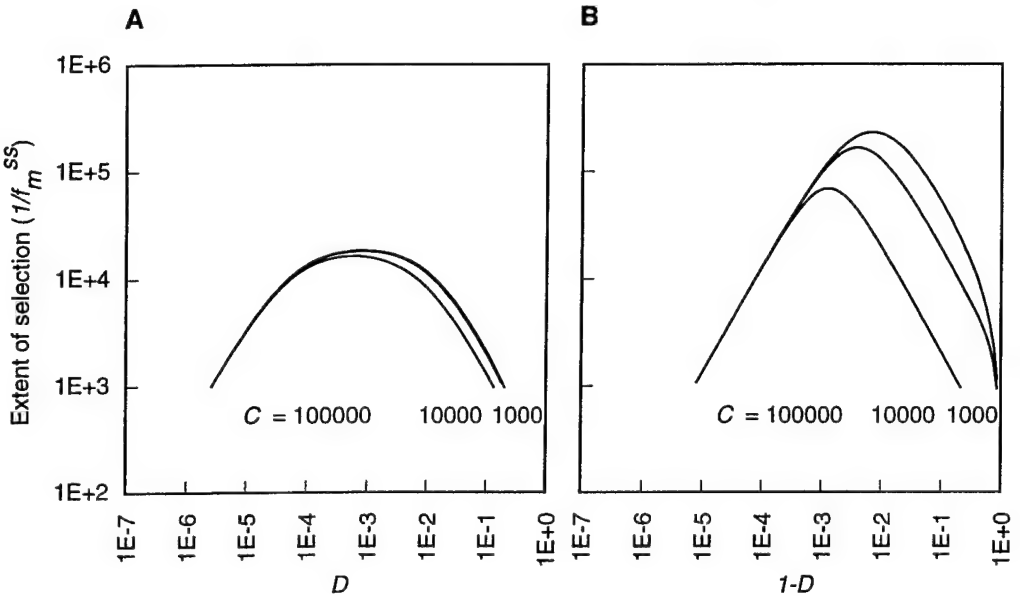


FIGURE 7.—The extent of selection as a function of demand for gene expression D and cycle time C. The extent of selection is given by the steady-state value of the ratio $X_w/(X_m + X_p)$ or $1/f_m^{ss}$. The parameters have the nominal values given in Table 1. (A) Negative mode of gene control. (B) Positive mode of gene control. See text for discussion.

ior of the populations is independent of the cycle time C. The time scale is actually discrete, given by values of nC, where n is the number of cycles. Thus, within a fixed time period, the same degree of enrichment can be achieved with either a large value for C and a small number n, or a small value for C and a larger number n. Note that the negative mode of gene control emerges more rapidly as demand for gene expression *increases*, whereas the positive mode of gene control emerges more rapidly as demand for gene expression *decreases* (Figure 6). The extent and rate of selection are examined in greater detail below.

Extent of selection: We define the *extent of selection* as the steady-state value of the ratio $X_w/(X_m + X_p)$, which is the inverse of the mutant fraction in the population $(1/f_m)$. Although there is selection for the wild-type regulatory mechanism throughout the region of overlap beneath the thresholds (*e.g.*, Figures 2A and 3A), the extent of the selection varies as a function of cycle time C and demand D. For a given value of C, the extent of selection reaches its maximum at a value of D that is roughly the geometric mean of its threshold values. With the nominal values for the parameters (Table 1), the results for the negative mode of gene control are as shown in Figure 7A; the results for the positive mode are similar to those for the negative mode, except that the allowable values for demand now occur in the high-demand region of the plot (Figure 7B). The maximum extent of selection for the positive mode of gene control is ~10-fold greater than that for the negative mode.

Rate of selection: The rate at which selection occurs is independent of cycle time. We define *response time* as the time required for the ratio $X_w/(X_m + X_p)$ to reach 99% of its steady-state value, starting from an initial state in which the numbers of the two types of mutants are equal and the ratio is equal to one-tenth of its steady-state value. Recall that the time points are given in units of nC, where C is the cycle time and n is the number of cycles. The same temporal behavior is obtained regardless of whether C is large (n small) or small (n large). However, the resolution is poorer for large values of C because the minimum value of n is one.

Like the extent of selection, the rate of selection is strongly dependent on the demand for gene expression. Although selection in the case of the negative mode can occur near the lower limit of allowable values for D, the response time is very long. Response time decreases in an inverse fashion as D increases, until a lower plateau is reached (Figure 8A). The break in the curve occurs at approximately the value of D that yields the maximum extent of selection (see Figure 7A). The minimum response time with the nominal values for the parameters is ~294,000 hr (~36 yr).

Similar results are found with the positive mode of gene control (Figure 8B), except that the long response times occur near the upper limit of allowable values for D. The response time decreases as D decreases until a minimum is reached, and then it increases. For the same extent of selection as the negative mode (18,400), the positive mode exhibits a faster response time (~17,000 *vs.* 294,000 hr); alternatively, for the same response time as the negative mode (294,000 hr), the positive mode exhibits a greater extent of selection (~214,000 *vs.* 18,400). Thus, it appears that the positive mode of gene control is capable of achieving greater extents of selection with faster response times than is the case for the negative mode of control.

Minimum cycle time: Estimates for the minimum cycle time of *E. coli* passing from one host to another can be obtained by combining the information in Figure 2A with the inverse relationship between C and D for the lactose operon of *E. coli*. (Recall from the ECOLOGY

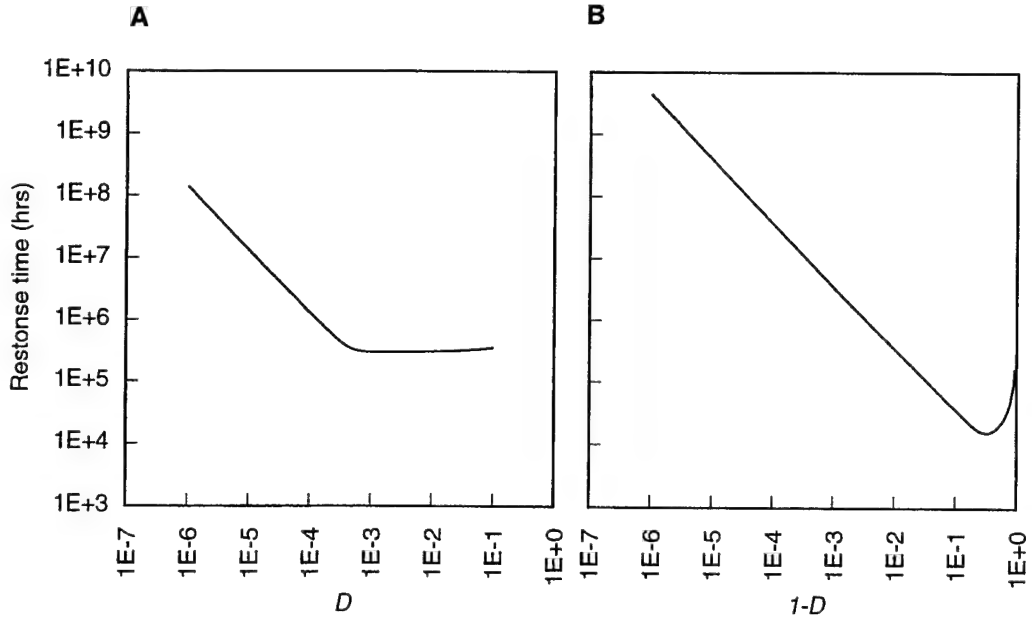


FIGURE 8.—The rate of selection as a function of demand for gene expression D. Response time is measured as the time required to achieve 99% of the steady-state value for the ratio $X_w/(X_m + X_p)$ starting from an initial state in which this ratio is initially one-tenth of the steady-state value and the mutants are equally abundant. The parameters have the nominal values given in Table 1. (A) Negative mode of gene control. (B) Positive mode of gene control. See text for discussion.

AND GENE EXPRESSION section that $C = 3/D$.) The intersection of this inverse relationship with the threshold for selection of the wild-type modulator (regulator) gives a value of $C_{min} = 26$ hr (Figure 9A). Another estimate of minimum cycle time can be obtained by combining the information in Figure 3A with the inverse relationship between C and $(1 - D)$ for the maltose operon of *E. coli*. [Recall from the ECOLOGY AND GENE EXPRESSION section that $C = 6/(1 - D)$.] The intersection of this inverse relationship with the threshold for selection of the wild-type modulator (regulator) gives a value of $C_{min} = 10$ hr (Figure 9B).

Maximum cycle time: Estimates for the maximum cycle time of *E. coli* passing from one host to another

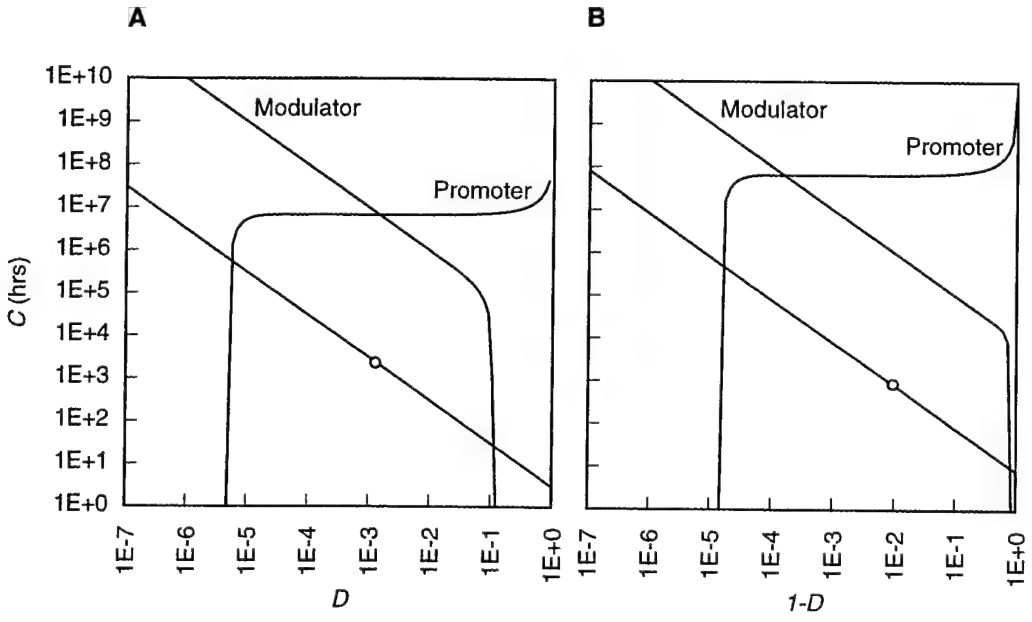


FIGURE 9.—Predicted values for cycle times C_{min}, C_{max}, and C_{op} based on the transit time through the proximal portions of the intestinal tract in humans. (A) Locus of C values for the negative mode of control is given by the $1/D$ relationship based on the period available for bacterial utilization of lactose (~3 hr). (B) Locus of C values for the positive mode of control is given by the $1/D$ relationship based on the period not available for utilization of maltose (~6 hr). The locus in each case intersects the threshold for selection of the wild-type modulator (regulator) at the value for C_{min} and the threshold for selection of the wild-type promoter at the value for C_{max}. The value of C_{op} is found on the locus at a value of D_{op}, which corresponds to the value of D that yields the optimum extent (Figure 7) and rate (Figure 8) of selection. See text for discussion.

also can be obtained by combining the information in Figure 2A with the inverse relationship between C and D for the lactose operon of *E. coli*. The intersection of the inverse relationship $C = 3/D$ with the threshold for selection of the wild-type promoter (Figure 9A) gives a value of $C_{max} = 580{,}000$ hr (~66 yr). Again, the data for the maltose operon in Figure 3A provide an alternative estimate. The intersection of the inverse relationship $C = 6/(1 - D)$ with the threshold for selection of the wild-type promoter (Figure 9B) gives a value of $C_{max} = 502{,}000$ hr (~57 yr).

Optimal cycle time: Although the estimates for minimum and maximum cycle time in the preceding sections are of some interest, perhaps the more relevant issue is the nominal value of the cycle time for *E. coli* in its natural environment. We argue that the most probable values for the cycle time will be those corresponding to the values for demand that lead to the optimal extent and rate of selection. For the negative mode, the optimum extent and rate of selection occur with a value of demand $D_{op} \approx 0.001$ (Figures 7A and 8A). Combining this optimum value for D with the inverse relationship $C = 3/D$ in Figure 9A for the lactose operon yields an estimate for the nominal value of the cycle time, namely, $C_{op} = 3000$ hr (~4 mon). For the positive mode, the optimum extent and rate of selection occur with a value of demand $1 - D_{op} \approx 0.01$ (Figures 7B and 8B). Combining this optimum value for $1 - D$ with the inverse relationship $C = 6/(1 - D)$ in Figure 9B for the maltose operon yields an estimate of $C_{op} = 800$ hr (~33 days).

DISCUSSION

The application of demand theory presented in this article provides an opportunity to test a number of the theory's quantitative implications. The results in Savageau (1998) led to the prediction of well-defined regions within the C *vs.* D plot where selection for the positive and negative modes of gene control is realizable. These regions allow one for the first time to specify precisely what is meant by high and low demand. With the nominal values for the parameters of the lactose and maltose operons in *E. coli*, selection of the negative mode of control requires a demand between 0.000005 and 0.1 (Figure 2A), whereas selection of the positive mode requires a demand between 0.2 and 0.999985 (Figure 3A). Furthermore, these regions were predicted to exhibit an inherent asymmetry with the positive mode having the larger region within which selection is realizable. This is clearly seen in the case of the lactose and maltose examples analyzed here (Figure 4).

Although the minimum and maximum values of demand are influenced by a number of parameters, by far the most influential parameter is σ, which reflects the reduction in growth rate when there is superfluous expression of a gene (Figure 5A and 5D). The nominal value for this parameter was set at 0.1%, on the basis of data for the lactose operon that suggest a value <1% for the reduction in growth rate of operator-constitutive mutants in a low-demand environment. In the case of the positive mode, the same value was used to characterize the reduction in growth rate of an up-promoter mutant in a low-demand environment. A 0.1% variation in σ yields a twofold change in the value of D_{max} for both the negative and positive mode (Table 2). The remaining parameters have much less influence on the limits of D; approximately one-half exhibit a nearly linear influence, whereas the other half have a negligible influence.

The ratio of mutation rate to selection coefficient is the most relevant determinant of the realizable region for selection. Indeed, if the target sizes for the various types of mutations are increased by an order of magnitude (*e.g.*, to match the footprint for binding a regulator protein to its modulator site on the DNA) at the same time the selection coefficients are increased by an order of magnitude, then the results are essentially unchanged (data not shown).

The results in Figure 5 suggest that the effect of transcription on mutation rate may be significant only if it reduces the mutation rate. The parameter ε, which represents this effect, has no influence on the selection of the wild-type promoter when there is a negative mode of control (Figure 5B). This is counter to the intuitive expectation that suggests a lower mutation rate would aid the selection of the wild-type promoter when it is not in use. The results in Figure 5A show that the parameter ε can represent an increased selection for the wild-type repressor-modulator interaction (increased D_{max}) if there is an increase in transcription-coupled repair (decrease in ε). In the case of the positive mode, ε has negligible influence on the selection of the wild-type activator-modulator interaction (Figure 5C), and its influence on the selection of the wild-type promoter (Figure 5D) would appear to be of little consequence because the threshold value of D_{max} in this case is already so high. Given the nominal values we have used for the parameters, ε does not seem to be highly significant, and similar effects can be achieved by varying other parameters; nevertheless, ε might still be important for selection under other conditions.

The equations that characterize the population dynamics of mutant and wild-type organisms (Equations 7–11 in Savageau 1998) led to the prediction that the extent of selection is a function of cycle time C and maximal at intermediate values of demand D, whereas the rate of selection is independent of cycle time C. Indeed, with the parameter values in Table 1, the extent of selection increases, reaches a maximum, and then declines as demand increases (Figure 7). As seen in Figure 8, the time required to reach full selection decreases until a minimum is reached with increasing demand (negative mode) or decreasing demand (positive mode). The combination of these results suggests that

the optimum extent and rate of selection occurs at around $D = 0.001$ for the negative mode and $1 - D = 0.01$ for the positive mode. In the case of the positive mode, this represents a choice of $1 - D$ that yields a rate of selection that is nearly equivalent to the optimum for the negative mode.

The quantitative theory reveals a number of new relationships involving cycle time that can be tested against experimental data in the case of the lactose and maltose operons of *E. coli*. The first such relationship provides an estimate for the minimum value of the cycle time C_{min}. We obtained values of 26 hr (Figure 9A) and 10 hr (Figure 9B), which is on the same order of magnitude as the 40 hr required on average for transit through the entire intestinal tract (CUMMINGS and WIGGINS 1976; GEAR *et al.* 1980; SAVAGEAU 1983). Under these circumstances, *E. coli* is simply passing through the intestinal tract without colonizing the colon. Clearly, the cycle time can be no shorter than this period.

The second relationship provides an estimate for the maximum value of the cycle time C_{max}. We have estimated this value to be ~580,000 hr (~66 yr) in the case of the lactose operon (Figure 9A) and 502,000 hr (~57 yr) in the case of the maltose operon (Figure 9B). These values for C_{max} are on the same order of magnitude as the 120-yr maximum for the life span of humans (HAYFLICK 1977). Clearly, the cycle time for *E. coli* can be no longer than the life time of the host because the bacteria will die with the host if they do not colonize a new host.

The final relationship provides an estimate for the optimum value of the cycle time C_{op}. The optimum extent and rate of selection determined for the lactose operon suggest a demand in the neighborhood of $D_{op} = 0.001$. This value of D, taken together with the relationship $D = 3/C$, predicts an optimum cycle time of $C_{op} = 3000$ hr (~4 mon). The corresponding estimate based on the maltose operon is $C_{op} = 800$ hr (~33 days). These predicted values for the cycle time of *E. coli* are comparable with the cycle times (recolonization rates) of months to years that have been observed in humans for resident strains of *E. coli* (SEARS *et al.* 1950; SEARS and BROWNLEE 1952; CAUGANT *et al.* 1981).

In summary, the quantitative development of demand theory presented in SAVAGEAU (1998) and applied here provides the first estimates for the minimum and maximum values of demand that are required for selection of the positive and negative modes of gene control. The specific application to the maltose and lactose operons of *E. coli* suggests that the positive and negative modes of control for these genes are subject to selection throughout the full range of cycle times that are possible for this microbe. Moreover, the cycle times predicted on the basis of the optimal extent and rate of selection are in agreement with the typical cycle times that have been observed experimentally.

I thank Drs. S. COOPER, R. G. FRETER, D. E. KIRSCHNER, J. V. NEEL, and M. S. SWANSON for critically reading the manuscript and two anonymous reviewers who made valuable suggestions for improving the parameter estimates. This work was supported in part by U.S. Public Health Service grant RO1-GM30054 from the National Institutes of Health and U.S. Department of Defense grant N00014-97-1-0364 from the Office of Naval Research.

LITERATURE CITED

BOND, J. H., and M. D. LEVITT, 1976 Quantitative measurement of lactose absorption. Gastroenterology **70:** 1058–1062.

CAUGANT, D. A., B. R., LEVIN and R. K. SELANDER, 1981 Genetic diversity and temporal variation in the *E. coli* population of a human host. Genetics **98:** 467–490.

COOKE, E. M., 1974 *Escherichia coli and Man.* Churchill Livingstone, London.

CUMMINGS, J. H., and H. S. WIGGINS, 1976 Transit through the gut measured by analysis of a single stool. Gut **17:** 219–223.

DAHLQVIST, A., 1961 The location of carbohydrases in the digestive tract of the pig. Biochem. J. **78:** 282–288.

DATTA, A., and S. JINKS-ROBERTSON, 1995 Association of increased spontaneous mutation rates with high levels of transcription in yeast. Science **268:** 1616–1619.

DRAKE, J. W., 1991 A constant rate of spontaneous mutation in DNA-based microbes. Proc. Natl. Acad. Sci. USA **88:** 7160–7164.

FRANCINO, M. P., L. CHAO and M. A. RILEY, 1996 Asymmetries generated by transcription-coupled repair in enterobacterial genes. Science **272:** 107–109.

FRETER, R., 1976 Factors controlling the composition of the intestinal micro-flora, pp. 109–120 in *Proceedings of the Microbial Aspects of Dental Caries*, special supplement, *Microbiology Abstracts*, Vol. 1, edited by H. M. STILES, W. J. LOESCHE and T. C. O'BRIEN. American Society for Microbiology, Washington, DC.

GEAR, J. S. S., A. J. M. BRODRIBB, A. WARE and J. T. MANN, 1980 Fiber and bowel transit times. Br. J. Nutr. **45:** 77–82.

HARLEY, C. B., and R. P. REYNOLDS, 1987 Analysis of *E. coli* promoter sequences. Nucleic Acids Res. **15:** 2343–2361.

HAYFLICK, L., 1977 The cellular basis for biological aging, pp. 159–186 in *Handbook of the Biology of Aging*, edited by C. E. FINCH and L. HAYFLICK. Van Nostrand Reinhold, New York.

KOCH, A. L., 1983 The protein burden of *lac* operon products. J. Mol. Evol. **19:** 455–462.

KOLDOVSKY, O., and F. CHYTIL, 1965 Postnatal development of β-galactosidase activity in the small intestine of the rat: effect of adrenalectomy and diet. Biochem. J. **94:** 266–270.

LECLERC, J. E., B. LI, W. L. PAYNE and T. A. CEBULA, 1996 High mutation frequencies among *Escherichia coli* and *Salmonella* pathogens. Science **274:** 1208–1211.

LISSER, S., and H. MARGALIT, 1993 Compilation of *E. coli* mRNA promoter sequences. Nucleic Acids Res. **21:** 1507–1516.

MAALØE, O., and N. O. KJELDGAARD, 1965 *Control of Macromolecular Synthesis.* Benjamin, New York.

MADSEN, J. L., 1992 Effects of gender, age, and body mass index on gastrointestinal transit times. Digest. Dis. Sci. **37:** 1548–1553.

MALAGELADA, J.-R., J. S. ROBERTSON, M. L. BROWN, M. REMINGTON, J. A. DUENES *et al.*, 1984 Intestinal transit of solid and liquid components of a meal in health. Gastroenterology **87:** 1255–1263.

MILLER, J. H., and W. S. REZNIKOFF, 1980 *The Operon.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

NOVICK, A., and M. WEINER, 1957 Enzyme induction as an all-or-none phenomenon. Proc. Natl. Acad. Sci. USA **43:** 553–566.

PRESS, W. H., B. P. FLANNERY, S. A. TEUKOLSKY and W. T. VETTERLING, 1988 *Numerical Recipes in C.* Cambridge University Press, New York.

REALLA, J. D., and J. COLLADO VIDES, 1996 Organization and function of transcription regulatory elements, pp. 1232–1245 in *Escherichia coli and Salmonella*, Vol. I, Ed. 2, edited by F. C. NEIDHARDT. ASM Press, Washington, DC.

ROSENSWEIG, N. S., and R. H. HERMAN, 1968 Control of jejunal sucrase and maltase activity by dietary sucrose or fructose in man: a model for the study of enzyme regulation in man. J. Clin. Invest. **47:** 2253–2262.

SAVAGEAU, M. A., 1977 Design of molecular control mechanisms and the demand for gene expression. Proc. Natl. Acad. Sci. USA **74:** 5647–5651.

SAVAGEAU, M. A., 1983 *Escherichia coli* habitats, cell types, and molecular mechanisms of gene control. Am. Naturalist **122:** 732–744.

SAVAGEAU, M. A., 1989 Are there rules governing patterns of gene regulation? pp. 42–66 in *Theoretical Biology—Epigenetic and Evolutionary Order*, edited by B. C. GOODWIN and P. T. SAUNDERS. Edinburgh University Press, Edinburgh.

SAVAGEAU, M. A., 1998 Demand theory of gene regulation. I. quantitative development of the theory. Genetics **149:** 1665–1676.

SCHNEIDER, T. D., G. D. STORMO, L. GOLD and A. EHRENFEUCHT, 1986 Information content of binding sites on nucleotide sequences. J. Mol. Biol. **188:** 415–431.

SCHWARTZ, M., 1987 The maltose regulon, pp. 1482–1502 in *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*, edited by F. C. NEIDHARDT. American Society for Microbiology, Washington, DC.

SEARS, H. I., and I. BROWNLEE, 1952 Further observations on the persistence of individual strains of *Escherichia coli* in the intestinal tract of man. J. Bacteriol. **63:** 47–57.

SEARS, H. I., I. BROWNLEE and J. K. UCHIYAMA, 1950 Persistence of individual strains of *E. coli* in the intestinal tract of man. J. Bacteriol. **59:** 293–301.

SHALLENBERGER, R. S., 1974 Occurrence of various sugars in foods, pp. 67–80 in *Sugars in Nutrition*, edited by H. L. SIPPLE and K. W. MCNUTT. Academic Press, New York.

SHIRAISHI, F., and M. A. SAVAGEAU, 1992 The tricarboxylic acid cycle in *Dictyostelium discoideum* II. Evaluation of model consistency and robustness. J. Biol. Chem. **267:** 22919–22925.

WALKER, D. G., 1968 Developmental aspects of carbohydrate metabolism, pp. 465–496 in *Carbohydrate Metabolism*, Vol. 1, edited by F. DICKENS, P. J. RANDLE and W. J. WHELAN. Academic Press, New York.

WIDDAS, W. F., 1971 The role of the intestine in sucrose absorption, pp. 155–171 in *Sugar*, edited by J. YUDKIN, J. EDELMAN and L. HOUGH. Butterworths, Woburn, MA.

Communicating editor: R. H. DAVIS

7 Reitzer, L. J. (1996) in *Escherichia coli* and *Salmonella*. Cellular and Molecular Biology (Neidhardt, F. C., Curtiss, III, R., Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B., Reznikoff, W. S., Riley, M., Schaechter, M. and Umbarger, H. E., eds.), pp. 391–407, American Society for Microbiology, Washington, DC
8 van Heeswijk, W. C., Stegeman, B., Hoving, S., Molenaar, D., Kahn, D. and Westerhoff, H. V. (1995) FEMS Microbiol. Lett. **132**, 153–157
9 van Heeswijk, W. C., Hoving, S., Molenaar, D., Stegeman, B., Kahn, D. and Westerhoff, H. V. (1996) Mol. Microbiol. **21**, 133–146
10 van Heeswijk, W. C. (1998) Ph.D. Thesis, University of Amsterdam, The Netherlands
11 Kacser, H. and Burns, J. A. (1983) Biochem. Soc. Trans. **7**, 1149–1150

Received 8 September 1998

Design of gene circuitry by natural selection: analysis of the lactose catabolic system in *Escherichia coli*

M. A. Savageau

Department of Microbiology and Immunology, The University of Michigan Medical School, Ann Arbor, MI 48109-0620, U.S.A.

Introduction

The need for a systems perspective in biology has never been more apparent than it is today. The accumulation of data concerning the basic determinants of biological systems was greatly accelerated when genetics, biochemistry and microbiology were fused to form molecular biology in the 1950s. However, this information has remained largely incomplete and fragmented. The new technologies that have grown out of the Human Genome Project have introduced a radically different approach, based on global measurements of the organism's phenotype. These global expression systems are producing a flood of data that must be related to the underlying molecular determinants. Without a quantitative systems theory with which to relate the information at these different levels of organization, our understanding will remain descriptive and lack predictive value.

Biochemical Systems Theory [1] is concerned with understanding integrated (systems level) behaviour in terms of the underlying (molecular level) determinants. This theory is based upon the power-law formalism [2], which provides a flexible, accurate and tractable mathematical representation for characterizing system components and their interactions. Biochemical Systems Theory provides methods of analysis that are capable of extracting information that is latent within the mathematical representation of the integrated system. Most importantly, Biochemical Systems Theory gives us a strategy for making well controlled comparisons that are at the heart of biological understanding in an evolutionary context. The primary aim of Biochemical Systems Theory is to elucidate the design principles that characterize intact biological systems.

Biochemical Systems Theory has been applied to several generic classes of metabolic pathways and gene circuits. Here I will examine three elements of design for a generic class of inducible gene circuits: threshold generation, gene coupling and mode of control. The principles that have been discovered in each case will be discussed in the context of the lactose (*lac*) system of *Escherichia coli*. I will finish with a few general conclusions that can be drawn from these examples.

Threshold generation

A sharp threshold for induction of a catabolic system will prevent premature induction of the catabolic machinery when there is an inadequate (subthreshold) supply of substrate in the organism's environment. Conversely, a sharp threshold will produce a highly responsive induction when the substrate supply is suprathreshold and sufficient not only to recoup the cost of synthesizing the catabolic machinery but also to provide the organism with excess carbon and energy for cellular growth and function [3]. Two alternative means for the generation of a sharp threshold are static and dynamic switches.

An example of a static switch is provided by an inducible catabolic system in which the substrate is the inducer that interacts with the regulator protein to produce a sigmoidal influence on the rate of mRNA transcription; all other steps in the system are operating in a first-order fashion.

If one increases the concentration of substrate (inducer) slowly from a low to a high value, gene expression will exhibit a low value that changes slowly, then accelerates through the mid-range values, and finally achieves a high value that again changes only slowly. If one decreases the concentration of substrate slowly from high to low values, the same values for gene expression will be retraced, but in the opposite order. Thus the process is completely reversible, with no memory for the past history of substrate concentration. The greater the degree of sigmoidicity (the larger the Hill number), the sharper the switch from low to high values for gene expression. (In the mathematical limit of an infinite Hill number, the sigmoidal characteristic approaches a step function with ony two values: a low value for expression when substrate concentrations are subthreshold, and a high value when they are suprathreshold.) This substrate induction model describes the *lac* system in a strain of *E. coli* that is freely permeable to the gratuitous inducer isopropyl β-D-thiogalactoside [4].

An example of a dynamic switch is provided by an inducible catabolic system that is similar to that described above, except that the product rather than the substrate of the induced enzymes is the inducer of gene expression. The properties of such a switch are quite different from those of the static switch. If one increases the concentration of substrate slowly from a low to a high value, the rate of mRNA synthesis will exhibit a low value that changes slowly, then jumps discontinuously to a higher value once a threshold concentration has been crossed, and finally remains at a high value that again changes only slowly. If one now decreases the concentration of substrate from high to low values, the rate of mRNA synthesis will exhibit a high value that changes slowly, then jumps discontinuously to a lower value once a second (lower) threshold concentration has been crossed, and finally remains at a low value that again changes only slowly. For substrate concentrations above the higher threshold and below the lower threshold, the system exhibits a single steady state. For substrate concentrations between these two threshold values, the system can be in one of two different stable steady states, depending upon which threshold was the last to be crossed. In this sense, the process is irreversible, with a memory of its past.

Induction of the *lac* system exhibits a dynamic switch when wild-type *E. coli* is exposed to lactose [5]. The conventional explanation suggests that transport of lactose into the cell is 'autocatalytic' in the following sense. The intracellular product of transport is the inducer, which causes further induction of the transport system, which in turn leads to a further increase in the concentration of inducer. Indeed, a dynamic switch can be realized by a product-inducible catabolic system, as described in the previous paragraph. However, the natural inducer of the *lac* operon is not a product of the system, but allolactose, an intermediate whose synthesis and degradation are both induced. When the position of the natural inducer is shifted from product to intermediate of the catabolic pathway, the same model that produced a dynamic switch now produces a static switch. Thus a dynamic switch is difficult to reconcile with the current model of the *lac* system, i.e. an unbranched pathway consisting of lactose transport (LacY) [6] and catabolism (LacZ) [7] that is subject to co-ordinate induction by an intermediate that has a sigmoidal influence on the control of transcription.

One way to rectify this inconsistency is to postulate a non-inducible alternative fate for the intermediate. If the alternative fate represents a minor contribution to the total degradation of intermediate (Figure 1A), then the system behaves essentially like an intermediate-induced system. As the concentration of substrate (extracellular lactose) is slowly increased, gene expression follows the static switch associated with the sigmoidal control of transcription (Figures 1B and 1C). If the alternative fate represents a major contribution to the total degradation of intermediate (Figure 1D), then the system behaves essentially like a product-induced system. Gene expression will then exhibit the following pattern as the concentration of substrate (extracellular lactose) is slowly increased (Figures 1E and 1F): (1) expression at first shows little change along the lower portion of the sigmoidal characteristic; (2) it jumps discontinuously to the upper portion as the upper threshold of substrate concentration is crossed (Figure 1E, curve a); and (3) it changes little thereafter along the upper portion of the sigmoidal characteristic. Gene expression follows a different pattern as the concentration of substrate (extracellular lactose) is slowly decreased: (1) expression at first shows little change along the upper portion of the sigmoidal characteristic; (2) it jumps discontinuously to the lower portion as the lower threshold of substrate concentration is crossed

(Figure 1E, curve b); and (3) it changes little thereafter along the lower portion of the sigmoidal characteristic. The system can remain in either the upper or the lower stable state when substrate concentrations have values between the lower and upper thresholds (Figure 1E, curve c). The net effect of substrate concentration on mRNA levels is shown in Figures 1(C) and 1(F).

The design principles for the generation of thresholds in these instances are revealed at the level of the integrated gene circuitry. Although an essential feature of both types of switches, i.e. the sigmoidal rate of mRNA synthesis, can be elucidated by studies of transcription initiation with isolated molecular components *in vitro*, this information alone is insufficient to distinguish between the two types. The effects of induction on the synthesis and degradation of inducer also play a critical role. This shows clearly the induction characteristics are a property of the intact system.

Coupling of elementary circuits

Genes interact to produce complex patterns of expression that define the phenotype of the organism. The best studied examples of gene interaction and the patterns of coupled expression that result are provided by elementary gene circuits in bacteria (Figure 2A). The expression of an effector gene and its cognate regulator gene

Figure 1

Thresholds for induction of gene expression

When the effect of induction on the synthesis of an inducer is less than or equal to that on degradation (A), gene expression exhibits a static switch (B, C), whereas when the effect on the synthesis of an inducer is greater than that on degradation (D), gene expression exhibits a dynamic switch (E, F). At any fixed concentration of substrate, the steady-state levels of mRNA and inducer are given by the intersection of the sigmoidal curve (the effect of inducer on mRNA levels) and the appropriate broken curve (the effect of mRNA on inducer levels) (B, E). The net effect of substrate concentration on mRNA levels is shown in (C) and (F). See the text for discussion.

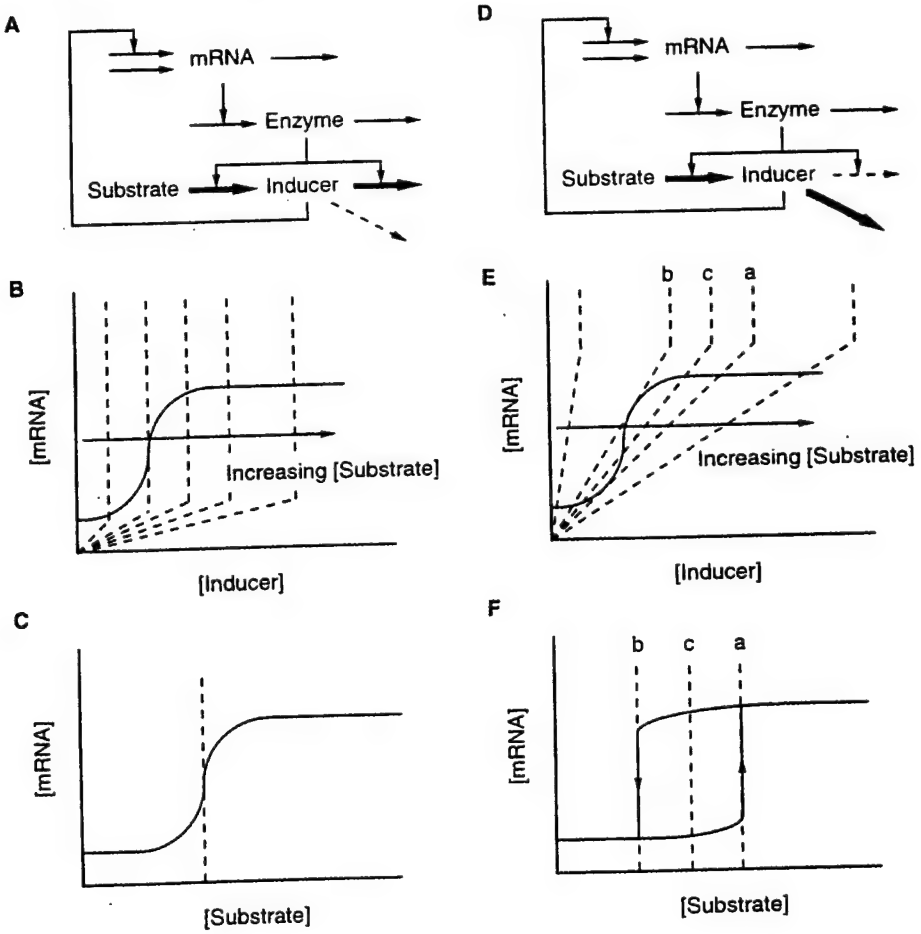
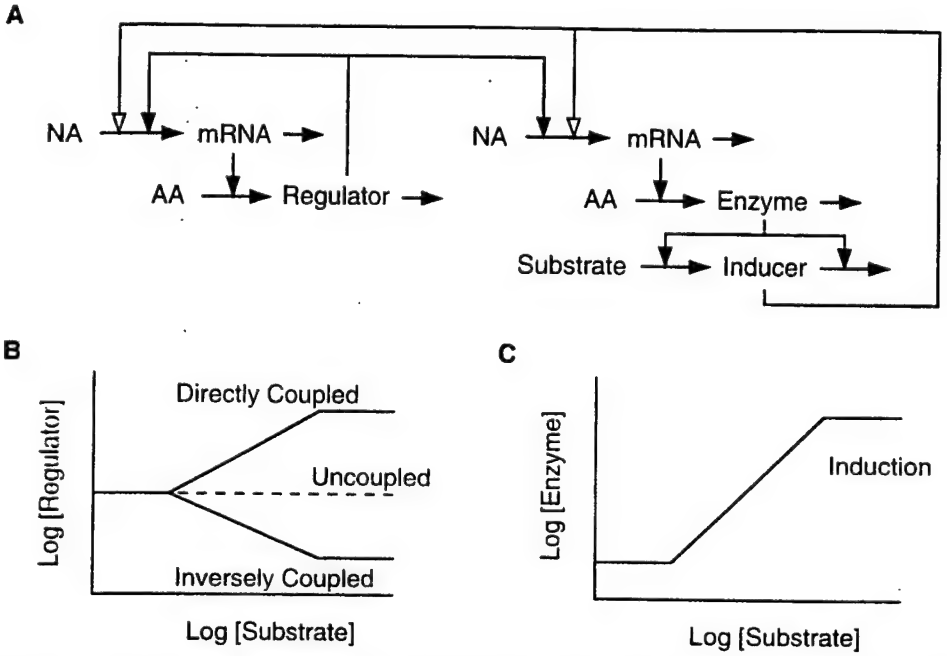


Figure 2

Coupling of expression in elementary gene circuits

(A) Linked circuits for regulator and effector genes. (B) Regulator expression can be directly coupled, uncoupled or inversely coupled with (C) effector expression. Abbreviations: NA, nucleic acids; AA, amino acids.



exhibits one of three distinct patterns of coupling. Induction of effector gene expression (Figure 2C) is accompanied by an increase, a decrease or no change in regulator gene expression (Figure 2B). Each of these three forms of coupling has been documented experimentally. However, until recently it was unclear whether the form of coupling follows some underlying rule or is the result of a frozen accident in the evolution of a given system. (For simplicity, in what follows, I shall ignore the inversely coupled patterns, since the decrease in regulator expression is small and in many cases difficult to distinguish from the case of uncoupled expression.)

There is now evidence to suggest that the pattern of coupling in these elementary gene circuits is governed by a simple design principle, which was discovered as follows. Biochemical Systems Theory was used to model a genetic system of elementary gene circuits capable of representing each of the distinct forms of coupling. The equations describing the system were solved both analytically and numerically over a wide range of parameter values. The resulting behaviour of the system with different parameter settings could then be classified on the basis of several criteria for functional effectiveness, including sharp threshold for induction, large logarithmic gain in product formation, robustness of the system in the face of parameter variation, regulator selectivity, system stability and temporal responsiveness. Finally, a strategy for making well-controlled comparisons among the results was followed, and a simple rule emerged. For systems that were well designed according to these *a priori* criteria, we discovered that if the regulator protein exerts negative control over effector gene expression, then expression of the regulator and effector genes is uncoupled when the capacity for induction is large and strongly coupled when the capacity for induction is small. Conversely, if the regulator protein exerts positive control over effector gene expression, then expression of the regulator and effector genes is uncoupled when the capacity for induction is small and strongly coupled when the capacity for induction is large [8,9].

The *lac* system, with negative control and a high capacity for induction, presents a superb example of this rule. Specific control of the *lac* system involves a negative regulator, the classical *lac* repressor. The capacity for β-galactosidase induction is large (approx. 1000-fold), whereas the capacity for *lac* repressor induction is essentially zero [10]. Thus expression of the regulator and effector genes in the *lac* system is

uncoupled, as expected for a system with a negative regulator and a large capacity for induction.

These predictions were tested against available experimental data for an additional 30 systems in which expression of both the regulator and effector genes had been measured, and agreement was found between the experimental measurements and the predicted pattern of coupling of gene expression [9]. This is an example of a design principle that is manifested at the level of integrated gene circuits. It could not have been discovered through the separate analysis of the individual circuits in isolation.

Molecular mode of control

The molecular mode of gene control exhibits a fundamental duality common to most, if not all, control systems. Control can be achieved either by removing a restraining element (the negative mode of control) or by providing a stimulatory element (the positive mode of control). Numerous examples of each mode have been documented in the literature. Although it was initially unclear whether rule or accident dictated the use of the positive or the negative mode in any particular system, Biochemical Systems Theory has led to the discovery of a surprisingly simple design principle that predicts the use of positive or negative control.

Biochemical Systems Theory was used to model systems exhibiting each of the alternative modes. An exhaustive analysis showed that, in most respects, the alternative systems behaved in an identical fashion. That is, systems with either the positive or the negative mode were functionally equivalent and could control gene expression equally well. However, their behaviour differed in diametrically opposed ways in response to genetic mutations that alter the components of the control system itself [3]. In response to damaging mutations, systems with the negative mode of control exhibited superfluous expression of the effector gene when it should be turned OFF, whereas systems with the positive mode of control failed to express the effector gene when it should be turned ON. A selectionist argument based on the population dynamics of mutant and wild-type organisms in different environments leads to the following principle: the positive mode of control will prevail when there is a high demand for effector gene expression in the organism's natural environment, whereas the negative mode will prevail when there is a low demand for expression [11,12].

This demand theory of gene regulation has now been developed in a more quantitative fashion [13], with key roles being played by two parameters. One is the average cycle time, C, taken for a gene to go from the OFF state to the ON state and back to the OFF state (Figure 3B). The other is the demand for expression, D, which is the fraction of the cycle time during which a gene is turned ON (Figure 3C). The solution of the population dynamic equations for mutant and wild-type organisms in alternative environments reveals a threshold for selection in the C against D plot. Selection of the wild-type control system can be realized only when values of C and D lie below the threshold (Figure 3D).

The regions of the C against D plot within which selection is possible differ for the alternative modes of control. The realizable region occurs at low values of D for the negative mode and at high values of D for the positive mode, and these regions exhibit an inherent asymmetry, with the realizable region for the positive mode being larger.

Application of this theory to the specific case of the lactose operon in *E. coli* cycling through the human intestinal tract (Figure 3A) yields several interesting predictions that relate to the host [14]. When the extremes of the realizable region intersect the inverse relationship between C and D, which is due to the fixed 3-h period of exposure to lactose during transit through the upper portion of the intestinal tract [15,16], one obtains predictions for the minimum and maximum cycle time (Figure 3D). The minimum value for the cycle time is approx. 26 h, which is roughly the time required for transit through the entire intestinal tract [17–19]. Under these conditions, *E. coli* passes through one intestinal tract after another as fast as possible without colonizing a single colon. The maximum value for the cycle time is approx. 580 000 h ($\sim$66 years), which is roughly equivalent to the life span of humans [20]. A longer cycle time would be impossible, because the colonizing bacteria would die with the host before recolonizing a new host.

This demand theory also makes predictions regarding the rate (Figure 3E) and extent (Figure 3F) of selection of the wild-type control system, and these exhibit optima that can be used to predict the nominal value for demand, which in turn leads to a prediction for the nominal cycle time (Figure 3D). The optimal extent and rate of selection determined for the lactose

operon suggest a demand (D) in the neighbourhood of 0.001. This value of D, taken together with the inverse relationship $D = 3/C$, predicts an optimal cycle time of 3000 h (~4 months), which is comparable with the cycle times (recolonization rates) that have been observed in humans for resident strains of *E coli* [21–23].

Demand theory provides an example of a biological design principle that only becomes manifest at the population level in the context of natural selection. Again, the demand principle could not have been discovered by examining molecular structures and interactions alone, nor could it have been discovered at the level of integrated gene circuits alone. The demand principle requires consideration of mutants and the environment external to the system.

Discussion

The details of Biochemical Systems Theory that are involved in the applications presented here are beyond the scope of this paper. However, a few general comments are in order.

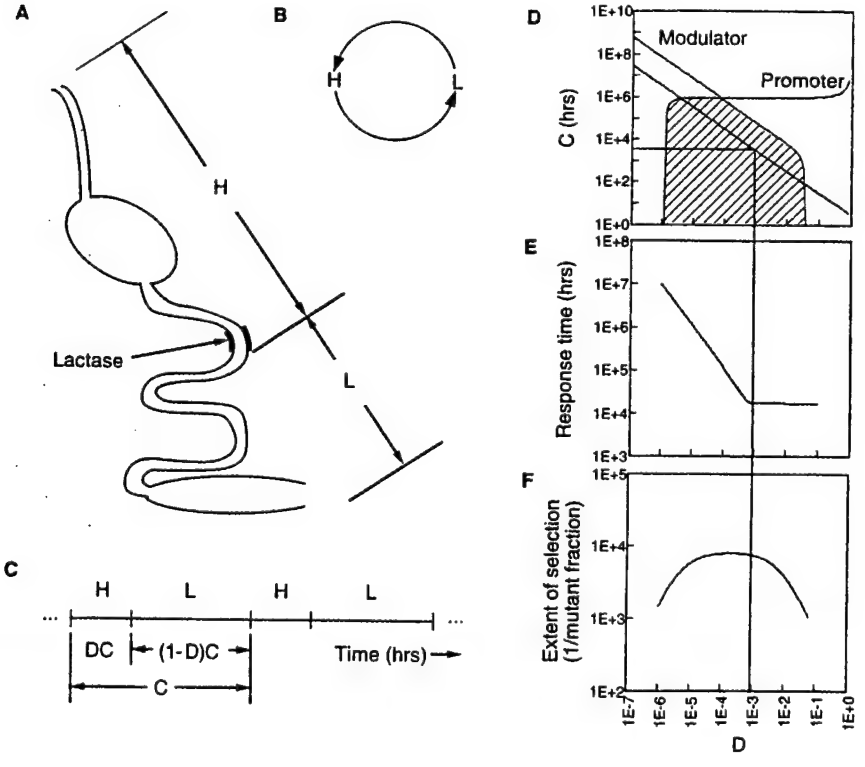
Biochemical Systems Theory has several general goals, including answers to the questions What, How and Why. What are the relevant components of the system under study? How do these components interact so as to produce the behaviour that is observed in the real system? Why is the system designed in this particular way and not some other? These and other goals are not unrelated, since the pursuit of one goal will often provide information that is needed in the pursuit of another. However, the primary goal of this theory is to discover the biological design principles that emerge at each level of organization in various generic classes of systems.

What is common to these successful explanations of design, and can this success be extended to more complex gene circuitry [24]? There are two aspects of these examples that seem most important. First, in each case one is able to identify a limited number of possible variations on a theme: static versus dynamic switches; coupled versus uncoupled circuits; positive versus negative modes of control. Even

Figure 3

Selection for the negative mode of control

A) Two regions of the human intestinal tract. (B) Cycling between the regions of high (H) and low (L) demand for lactose expression in *E. coli*. (C) Definition of average cycle time (C) and average demand (D) for gene expression. (D) Cycle time for selection, (E) rate of selection and (F) extent of selection, as a function of demand. See the text for discussion. 1E indicates exponential, i.e. $1E+2 = 10^2$; $1E-7 = 10^{-7}$; $1E+0 = 10^0$ etc.



though these variations were not understood initially, there was the prospect that a simple rule could be found, if indeed there was one, because of the limited number of variations in design that had to be analysed. Secondly, by an appropriate choice of organizational level and type of representation, one could obtain simple equations whose structure was amenable to qualitative analysis (and to exhaustive numerical analysis when necessary), and this leads to general results that are independent of specific values for the parameters. This is important, because many of the parameter values for any system will always be unknown. The success of Biochemical Systems Theory in elucidating general rules that are consistent with patterns found in nature is indicative of the power of this approach.

It is clear that the elucidation of design principles and the compilation of molecular detail lead to very different kinds of understanding. This is seen in the examples considered in the previous sections. Numerous experimental studies of gene circuitry over the past 30 years have produced an abundance of molecular descriptions and have documented the existence of each of the design features considered here. However, these experimental results provided no insight into the underlying principles that govern these designs. A focus on the kind of understanding that emerges from knowledge of the underlying design principles will become increasingly important in biochemistry, not only for advancing our research programmes, but also, and perhaps more importantly, for instructing the next generation of students. The current approach is overwhelming students with enormous amounts of information to memorize and providing less and less motivation for them to do so. A focus on design principles would provide deeper understanding, diminish the burden of memorization, and integrate their understanding into a broader and more meaningful context.

This work was supported in part by U.S. Public Health Service grant RO1-GM30054 from the National Institutes of Health and by U.S. Department of Defense grant N00014-97-1-0364 from the Office of Naval Research. I thank Grenmarie Agresar and Stephen Cooper for critically reading the manuscript.

1 Savageau, M. A. (1996) in Integrative Approaches to Molecular Biology (Collado-Vides, J., Magasanik, B. and Smith, T. F., eds.), pp. 115–146, MIT Press, Cambridge, MA
2 Savageau, M. A. (1998) BioSystems **47**, 9–36
3 Savageau, M. A. (1976) Biochemical Systems Analysis: A Study of Function and Design in Molecular Biology, Addison–Wesley, Reading, MA
4 Sadler, J. R. and Novick, A. (1965) J. Mol. Biol. **12**, 305–327
5 Novick, A. and Weiner, M. (1957) Proc. Natl. Acad. Sci. U.S.A. **43**, 553–566
6 Lolkema, J., Carrasco, N. and Kaback, H. (1991) Biochemistry **30**, 1284–1290
7 Huber, R. E., Wallenfels, K. and Kurz, G. (1975) Can. J. Biochem. **53**, 1035–1038
8 Hlavacek, W. S. and Savageau, M. A. (1995) J. Mol. Biol. **248**, 739–755
9 Hlavacek, W. S. and Savageau, M. A. (1996) J. Mol. Biol. **255**, 121–139
10 Beckwith, J. (1987) in *Escherichia coli* and *Salmonella typhimurium*: Cellular and Molecular Biology (Neidhardt, F. C., ed.), pp. 1444–1452, American Society for Microbiology, Washington, DC
11 Savageau, M. A. (1977) Proc. Natl. Acad. Sci. U.S.A. **74**, 5647–5651
12 Savageau, M. A. (1989) in Theoretical Biology: Epigenetic and Evolutionary Order (Goodwin, B. C. and Saunders, P. T., eds.), pp. 42–66, Edinburgh University Press, Edinburgh
13 Savageau, M. A. (1998) Genetics **149**, 1665–1676
14 Savageau, M. A. (1998) Genetics **149**, 1677–1691
15 Bond, J. H. and Levitt, M. D. (1976) Gastroenterology **70**, 1058–1062
16 Malagelada, J.-R., Robertson, J. S., Brown, M. L., Remington, M., Duenes, J. A., Thomforde, G. M. and Carryer, P. W. (1984) Gastroenterology **87**, 1255–1263
17 Cummings, J. H. and Wiggins, H. S. (1976) Gut **17**, 219–223
18 Gear, J. S. S., Brodribb, A. J. M., Ware, A. and Mann, J. T. (1980) Br. J. Nutr. **45**, 77–82
19 Savageau, M. A. (1983) Am. Nat. **122**, 732–744
20 Hayflick, L. (1977) in Handbook of the Biology of Aging (Finch, C. E. and Hayflick, L., eds.), pp. 159–186, Van Nostrand Reinhold, New York
21 Sears, H. I., Brownlee, I. and Uchiyama, J. K. (1950) J. Bacteriol. **59**, 293–301
22 Sears, H. I. and Brownlee, I. (1952) J. Bacteriol. **63**, 47–57
23 Caugant, D. A., Levin, B. R. and Selander, R. K. (1981) Genetics **98**, 467–490
24 Neidhardt, F. C. and Savageau, M. A. (1996) in *Escherichia coli* and *Salmonella typhimurium*: Cellular and Molecular Biology, 2nd edn. (Neidhardt, F. C., ed.), vol. 1, pp. 1310–1324, American Society for Microbiology, Washington, DC

Received 21 August 1998

Proceedings

Editors:
E. Bornberg-Bauer, A. De Beuckelaer,
U. Kummer, U. Rost

Workshop on Computation of Biochemical Pathways and Genetic Networks

Villa Bosch, Heidelberg
August 12–13, 1999

Evolution and Regulation of Gene Circuitry: Functional Genomics of the Lactose Pathway in E. Coli

Michael A. Savageau
Department of Microbiology and Immunology
The University of Michigan Medical School
Ann Arbor, Michigan 48109-0620
USA

Abstract

The physical basis for biological complexity is context-dependent expression of the organism's genome. The context is provided by the life cycle of the organism; the molecular mechanisms of gene regulation interpret that context. The relationship between these two different hierarchical levels of organization – genotype and phenotype – has traditionally been approached from the bottom-up perspective. Regulation of many gene systems has been studied in detail, and the results have revealed an enormous diversity of molecular elements and circuits. We are just beginning to understand the functional implications of such variations in design and to grasp the factors that have influenced their evolution. The relationship between genotype and phenotype can now be approached from the top-down perspective as well. The new technologies that have grown out of the Human Genome Project have introduced a radically different approach based on global measurements of the organism's phenotype. However, there are theoretical limits regarding the extent to which knowledge of the underlying mechanisms can be determined solely on the basis of systemic measurements. Success in relating genotype to phenotype will ultimately require a combination of both the top-down and bottom-up approaches. It also will require an appropriate systems theory for relating the information at these two levels of organization. Without a quantitative systems theory to relate the information at these different levels of organization our understanding will remain descriptive and lack predictive value. I will describe recent work on a quantitative theory that relates molecular mechanisms of gene control to the organism's physiological behavior in its natural environment. When applied to the lactose operon of Escherichia coli in the human intestine, the theory predicts selection for the correct mode of gene control. It also makes surprising predictions concerning the organism's phenotype and habitat.

Introduction

What is the function of regulatory gene circuitry? The superficial answer is fairly obvious. The genotype is determined by the information encoded in the DNA sequence, the phenotype is determined by the context-dependent expression of the genome, and the regulatory circuitry interprets the context and orchestrates the expression. However, a deeper

look reveals a hierarchy of mechanisms linking the genotype to the phenotype. At each level of this hierarchy there are various designs that are poorly understood. While some of the differences may be attributed to historical accidents with no attendant functional implications, others are governed by rules that result from natural selection [1].

The current goal of functional genomics is identifying the function of a gene product that is encoded in a particular string of nucleotides. Given a sequence can one predict that it codes for a dehydrogenase? From the particular combination of domains can one further deduce that it is in fact an homoserine dehydrogenase? These goals for relating genotype to phenotype, though still fairly modest, have yet to be fully realized. Nevertheless, it is instructive to look beyond the current status and examine the prospects of achieving the ultimate goal of functional genomics, which is to relate the nucleotide sequence of the genome to the expression of function in time and space given an appropriate context.

In this paper I will briefly review various hierarchical levels of organization from DNA sequence to environmental context, summarize results from three different analyses that connect the several levels of organization, and then show how these results provide a self-consistent relationship between genotype and phenotype of the organism in its natural environment in the case of a simple well-studied system, the lactose operon of Escherichia coli. The results in this case suggest that reaching the ultimate goal of functional genomics will involve more than recognizing complex patterns in the DNA sequence. It will require methods to elucidate and to quantify the complex web of interactions that link the numerous hierarchical levels of organization from DNA sequence to integrated behavior of the intact organism in its environment. Understanding the evolution and regulation of gene circuitry is at the heart of the matter.

Hierarchies from Genotype to Phenotype

The behavior of an intact biological system can seldom be related directly to its underlying molecular determinants. There are several different levels of hierarchical organization that intervene – genome sequence, transcriptional unit, mode of control, logic of control, expression cascade, connectivity, and environmental context.

The genome sequence consists of the four bases – A, T, G, and C – strung together in seemingly endless variation. However, genetic analysis has shown that a basic grammar defines how this alphabet is organized into meaningful units of transcription. These units consist of structural genes bounded by initiation and termination sites with a number of adjacent regulatory sites capable of binding specific transcription factors that modulate the rate of transcription initiation or termination.

The molecular mode of control by which individual regulator proteins affect the transcriptional process can be either positive or negative. For example, induction of gene expression can be achieved either by supplying the activator of a quiescent process or by removing the repressor of a constitutively active process. A particular set of such regulators, each with their own mode of action, can act in a combinatorial fashion on a single transcriptional unit to achieve control according to a specific logical function.

The transcription of a gene is only the first of many steps in a cascade of information flow from DNA to RNA to protein to metabolite that constitutes expression of a gene. These expression cascades are interconnected because the products of one cascade act as regulators of other cascades. If fact, it is the topological connectivity of the elementary expression cascades that constitutes the regulatory gene circuitry of the cell.

The environment provides the context that must be interpreted by the circuitry in order to produce a pattern of gene expression conducive to the organism's survival. These several levels of organization have been abstracted from what is known of gene expression in both prokaryotes and eukaryotes, but for testing our predictions we shall henceforth focus our attention on the lac operon of E. coli.

Lac Expression and Underlying Circuitry

It has long been known that expression of the lac operon can be induced by growth on lactose as a carbon and energy source [2]. However, the elucidation of the underlying circuitry involved the experimental exploitation of mutants and gratuitous inducers that allowed specific aspects of the regulatory mechanisms to be identified and characterized [3]. For example, the first evidence of cooperativity in the action of lac repressor was uncovered through the use of a non-metabolizable inducer and a transport mutant that together allowed the intracellular concentration of inducer to be specified experimentally via the extracellular environment [4].

Induction of the wild-type operon by gratuitous inducers provided evidence for a dynamic switch that exhibits hysteresis as shown in Figure 1B. When there is a low level of expression, substrate concentration must be increased above level a before there is an abrupt switch to a high level of expression; when there is a high level of expression, substrate concentration must be decreased below level b before there is an abrupt switch to a low level of expression. Thus, at intermediate concentrations of substrate, such as level c, the level of expression can be either high or low, depending upon the past history of the substrate concentration. The original explanation for this behavior focused on the positive feedback resulting from induction of the lac-encoded transport system [5]. This continued to be the accepted explanation even as the details of lac circuitry were characterized and a kinetic model developed.

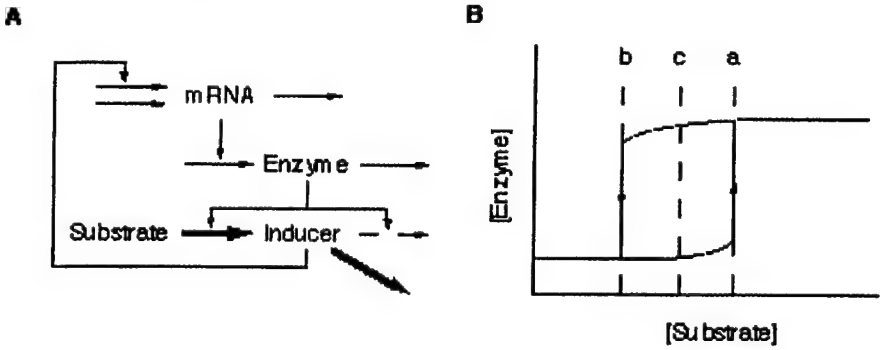


Figure 1: Alternative to the current kinetic model of the lac operon. (A) Schematic diagram showing the extra fate for the inducer allolactose. (B) Steady-state induction characteristic of the model in panel A exhibits hysteresis.

Recent analysis has demonstrated that the current kinetic model of the wild-type operon is capable of the dynamic switch that generates hysteresis in response to gratuitous inducers but not in response to the natural substrate lactose [6]. It is experimentally more difficult to obtain the steady-state induction characteristic with the natural substrate, so it is perhaps not surprising that such experiments are difficult to find in the literature. Special care must be taken to design experiments that are sufficiently long-term to insure a steady state and with sufficiently low cell densities to ensure negligible consumption of substrate. Carefully executed experiments of this kind have demonstrated an hysteretic response to lactose [7]. These two facts – experimental demonstration of hysteresis and mathematical demonstration that the current model is incapable of such behavior – indicates that the current model of the lac operon is inadequate.

We have proposed an alternative model (Figure 1A) in which the inducer has an additional fate, not inducible by allolactose, that accounts for the dynamic switch in

expression [6] when the additional flux is greater than a well-defined minimum [8]. We also have found independent experimental evidence in the literature that lends support to this model. The meaning of these results will become clear later in the context of results from other analyses.

Molecular Mode of Control and Demand for Lac Gene Expression

The conclusions from the analysis described in the previous section link knowledge of the organism's physiology to knowledge of the underling gene circuitry. The analysis in this section will link knowledge of the organism's genome and molecular mode of control to knowledge of the demand for gene expression during the organism's life cycle.

The life cycle of E. coli involves sequential colonization of new host organisms [9], which means repeated cycling between two different environments (Figure 2A and 2B). In one, the upper portion of the host's intestinal track, the microbe is exposed to the substrate lactose and there is a high demand for expression of the lac operon, and in the other, the lower portion of the intestinal track and the environment external to the host, the microbe is not exposed to lactose and there is a low demand for lac expression. The average time to complete a cycle through these two environments is defined as the cycle time, C, and the average fraction of the cycle time spent in the high-demand environment is defined as the demand for gene expression, D (Figure 2C).

The implications for gene expression of mutant and wild-type operons in the high- and low-demand environments are as follows. The wild-type functions by turning on expression in the high-demand environment and turning off expression in the low demand environment. The mutant with a defective promoter is unable to turn on expression in either environment. The mutant with a defective modulator (which can also stand in for a mutant with a defective regulator protein) is unable to turn off expression regardless of the environment. The double mutant with defects in both promoter and modulator behaves like the promoter mutant and is unable to turn on expression in either environment.

The mutation rates between these populations depend on the mutation rate per base and on the size of the relevant target sequence. The population sizes also are dependent upon selection.

There is selection against mutants of the modulator type in the low-demand environment because there is superfluous expression of an unneeded function. There is selection against mutants of the promoter type in the high-demand environment because expression of a needed function is lacking.

Solution of the dynamic equations for each of the populations cycling through the two environments yields equations in C and D for the threshold, extent, and rate of selection for the wild-type control mechanism [10]. The threshold for selection is shown by the shaded region in Figure 2D; only systems with values of C and D that fall within this region are capable of being selected. The rate and extent of selection shown in Figure 2E and 2F exhibit optimum values for a specific value of D.

Application of this theory to the lac operon of E. coli yields several new and provocative predictions that relate genotype to phenotype [11].

The straight line in Figure 2D represents the inverse relationship C=3D3/D that results from fixing the time of exposure to lactose at 3 hours, which is the clinically determined value for humans [12, 13]. The intersections of this line with the two thresholds for selection provide lower and upper bounds on the cycle time. The lower bound is approximately 24 hours, which is about as fast as the microbe can cycle through the intestinal track without colonization [14, 15, 16]. The upper bound is approximately 70 years, which is the longest period of colonization without cycling and corresponds favorably with the maximum life span of the host [17]. The optimum value for the cycle time, corresponding to the optimum

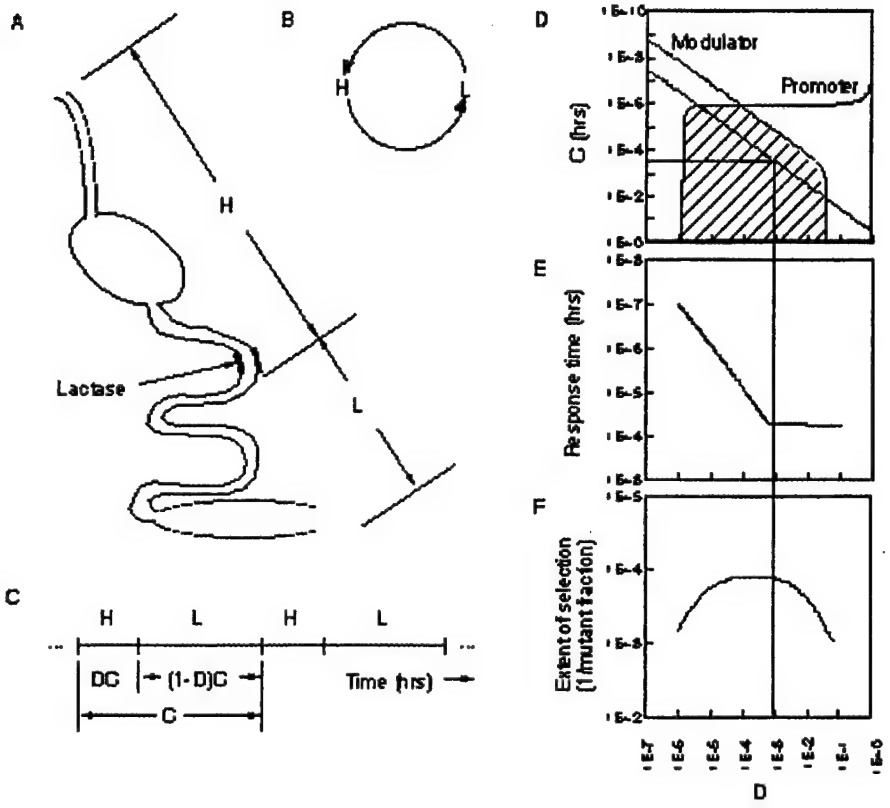


Figure 2: Life cycle and demand for gene expression. (A) Schematic diagram of the upper (high demand) and lower (low demand) portions of the human intestinal track. (B) Life cycle consists of repeated cycling between high- and low-demand environments. (C) Definition of cycle time C and demand for gene expression D. (D) Region in the C vs. D plot for which selection of the wild-type control mechanism is possible. (E) Rate of selection as a function of demand. (F) Extent of selection as a function of demand. See text for discussion.

value for demand (from Figures 2E and 2F), is approximately four months; this value is comparable with the average rate of recolonization measured in humans [18, 19, 20].

Logic and Phasing of Lac Control

The analysis in the previous section assumed that when E. coli was growing on lactose there was no other more preferred carbon source present. Thus, the positive CAP-cAMP regulator [21] was always present, and we could then concentrate on the conditions for selection of the specific control by Lac repressor. This was a simplifying assumption; in the general situation, both the specific control by Lac repressor and the global control by CAP-cAMP activator must be taken into consideration. The analysis becomes more complex, but it follows closely the outline of the simpler case in the previous section.

By extension of the definition for demand D, given in the pervious section, one can define a period of demand for the absence of repressor G, a period of demand for the presence of activator E, and a phase relationship between these two periods of demand F. By extension of the analysis in the previous section, solution of the dynamic equations for the wild-type and each of the mutant populations cycling through the two environments yields equations in C, G, E, and F for the threshold, extent, and rate of selection for the wild-type control mechanism [22].

The threshold for selection is now an envelope surrounding a "mound" in four- dimensional space with cycle time C as a function of the three parameters G, E, and F; only systems with values that fall within this envelope are capable of being selected. The rate and extent of selection exhibit optimum values as before, but these now occur with a specific combination of values for G, E, and F. The values of G, E, and F that yield the optima represent a small period when repressor is absent, an even smaller period when activator is present, and a large phase period between them. The period when repressor is absent corresponds to the period of exposure to lactose. Within this period there is a shorter period when activator is absent; this corresponds to the presence of a more preferred carbon source that lowers the level of cAMP.

These relationships can be interpreted in terms of exposure to lactose, exposure to glucose, and expression of the lac operon as shown in Figure 3. The initial exposure to lactose leads to an accumulation of the natural inducer allolactose and hence to induced expression of β-galactosidase. The result is a greatly increased synthesis of the products allolactose, glucose and galactose. These products are preferentially excreted back into the extracellular environment, in agreement with our proposed model for hysteretic expression of the lac operon. The extracellular glucose causes catabolite repression and lactose exclusion, thereby initiating a period of growth on glucose. During this period the activator CAP-cAMP is absent, transcription of the lac operon ceases and the concentration of β-galactosidase is diluted by growth, and lactose is spared. Eventually, glucose becomes depleted, the residual lactose causes a diminished secondary induction of β-galactosidase, and the microbe enters the low-demand environment as the lactose is exhausted.

Discussion

The results of the three different analyses described in the preceding sections are remarkably self-consistent and supported by a diverse set of independent experimental observations.

First, the analysis of lac circuitry showed that the conventional kinetic model is inadequate in that it is incapable of producing all-or-none expression of the lac operon. The new model we have proposed includes a non-inducible pathway for removal of inducer. There is indeed independent experimental evidence to support this model [23]. The natural inducer allolactose as well as the products of its hydrolysis are rapidly excreted from

the cell. Moreover, studies using mutants have shown that internally generated glucose is incapable of being used efficiently by the cell, whereas the excreted glucose, which is phosphorylated as it is transported back into the cell, is used with high efficiency. From these results one can see that the pathway for utilization of lactose involves induction of transport and catabolism of lactose, efflux of the glucose and galactose produced, transport of the external glucose and galactose back into the cell, and finally entry into the pathways of intermediary metabolism. The catabolite repression caused by glucose acts as a negative feedback mechanism to moderate the overall rate of lactose utilization.

Second, the quantitative version of demand theory integrates information at the level of DNA (mutation rate, effective target sizes for mutation of regulatory proteins, promoter sites, and modulator sites), physiology (selection coefficients for superfluous expression of an unneeded function and for lack of expression of an essential function), and ecology (environmental context and life cycle) and makes rather surprising predictions connected to the intestinal physiology and life span of the host and to the rate for recolonizing the host. There is independent experimental data to support each of these predictions.

Finally, when the logic of combined control by CAP-cAMP activator and Lac repressor was analyzed, we found an optimum set of values not only for the exposure to lactose, but also for the exposure to glucose and for the relative phasing between the periods of exposure. The phasing predicted is consistent with a self-generated glucose effect produced by catabolism of lactose and excretion of glucose. These results are supported by the same experimental data noted above in connection with the hysteretic expression of the lac operon.

The results from all three analyses fit together nicely. In the end, we are able to relate information in the nucleotide sequence of the lac operon to its specific pattern of expression in time and space. In the process we have made use of information at several other levels of organization, including important information about the host that provides the ecological niche for the microbe. While much of the necessary information could in principle be deduced from the underlying sequence, some would still require a systemic integration at the level of the intact organism and its environment. From this perspective it is clear that a completely reductionist deduction of function solely from information in the DNA sequence will be unattainable in most cases. Aside from those few organisms that have a relatively self-contained developmental program, functional genomics will ultimately be concerned with the genomes of multiple organisms undergoing mutual interaction and coevolution.

Acknowledgments

This work was supported in part by U.S. Public Health Service Grant RO1-GM30054 from the National Institutes of Health, and U.S. Department of Defense Grant N00014-97-1-0364 from the Office of Naval Research.

Bibliography

[1] Savageau, M.A. (1989) in Theoretical Biology – Epigenetic and Evolutionary Order, edited by B. C. Goodwin and P. T. Saunders. Edinburgh University Press, Edinburgh, pp. 42-66.

[2] Monod, J. and Audureau, A. (1947) Ann. Inst. Pasteur 72, 868-878.

[3] Beckwith, J.R. and Zipser, E., Eds. (1970) The Lactose Operon, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

[4] Sadler, J.R. and Novick, A. (1965) J. Mol. Biol. 12, 305-327.

[5] Novick, A. and Weiner, M. (1957) Proc. Nat. Acad. Sci. USA 43, 553-566.

[6] Savageau, M.A. (1999) Biochem. Soc. Trans. 27, 264-270.

[7] Ninfa, A., Eriksson, C. and Savageau, M.A. (in preparation).

[8] Eriksson, C., Ninfa, A. and Savageau, M.A. (in preparation).

[9] Cooke, E.M. (1974) Escherichia coli and man. Churchill Livingstone, London.

[10] Savageau, M.A. (1998) Genetics 149, 1665-1676.

[11] Savageau, M.A. (1998) Genetics 149, 1677-1691.

[12] Bond, J.H. and Levitt, M.D. (1976) Gastroenterology 70, 1058-1062.

[13] Malagelada, J.-R., Robertson, J.S., Brown, M.L., Remington, M., Duenew, J.A., Thomforde, G. M. and Carryer. P. W. (1984) Gastroenterology 87, 1255-1263.

[14] Cummings, J.H. and Wiggins, H.S. (1976) Gut 17, 219-223.

[15] Gear, J.S.S., Brodribb, A.J.M., Ware, A. and Mann, J.T. (1980) Br. J. Nutr. 45, 77-82.

[16] Savageau, M.A. (1983) Am. Naturalist 122, 732-744.

[17] Hayflick, L. (1977) in Handbook of the Biology of Aging, edited by C.E. Finch and L. Hayflick. VanNostrand Reinhold, N.Y., pp. 159-186.

[18] Sears, H.I., Brownlee, I. and Uchiyama, J.K. (1950) J. Bacteriol. 59, 293-301.

[19] Sears, H.I. and Brownlee, I. (1952) J. Bacteriol. 63, 47-57.

[20] Caugant, D.A., Levin, B.R. and Selander, R.K. (1981) Genetics 98, 467-490.

[21] Garges, S. (1994) in Transcription: Mechanisms and Regulation, edited by R.C. Conaway and J.W. Conaway. Raven Press, N.Y., pp. 343-352.

[22] Savageau, M.A. (in preparation).

[23] Huber, R.E. and Hurlburt, K.L. (1984) Can. J. Microbiol. 30, 411-418.

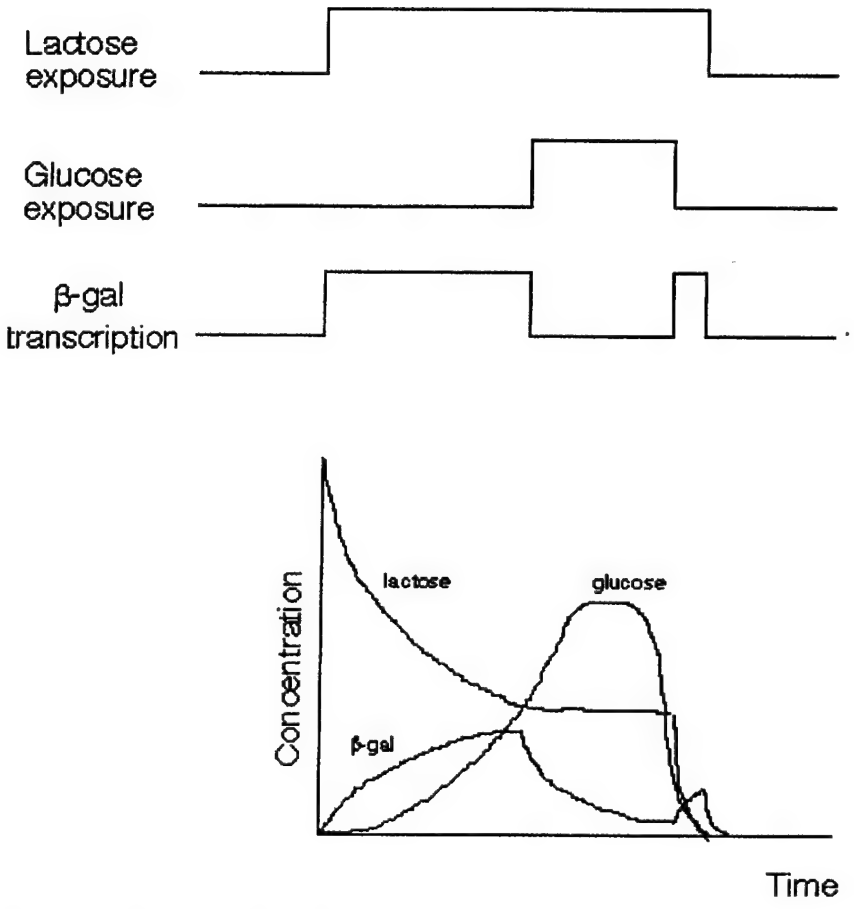


Figure 3: Schematic interpretation of optimal phasing of β-galactosidase expression in terms of exposure to lactose and glucose. See text for discussion.

BIOINFORMATICS

Vol. 16 no. 6 2000
Pages 527–533

Comparing systemic properties of ensembles of biological networks by graphical and statistical methods

Rui Alves [1, 2, 3] and Michael A. Savageau [1,*]

[1]Department of Microbiology and Immunology, University of Michigan Medical School, 5641 Medical Science Building II Ann Arbor, MI 48109-0620, USA, [2]Grupo de Bioquimica e Biologia Teoricas, Instituto Rocha Cabral, Calçada Bento da Rocha Cabral 14, 1250 Lisboa, Portugal and [3]Programa Gulbenkian de Doutoramentos em Biologia e Medicina, Departamento de Ensino, Instituto Gulbenkian de Ciência, Rua da Quinta Grande 6, 1800 Oeiras, Portugal

Received on July 28, 1999; revised on November 11, 1999; accepted on December 3, 1999

Abstract

Motivation: *When dealing with questions that concern a general class of models for biological networks, large numbers of distinct models within the class can be grouped into an ensemble that gives a statistical view of the properties for the general class. Comparing properties of different ensembles through the use of point measures (e.g. medians, standard deviations, correlation coefficients) can mask inhomogeneities in the correlations between properties. We are therefore motivated to develop strategies that allow these inhomogeneities to be more easily detected.*

Results: *Methods are described for constructing ensembles of models within the context of a Mathematically Controlled Comparison. A Density of Ratios Plot for a given systemic property is then defined as follows: the y axis represents the value of the systemic property in a reference model divided by the value in the alternative model, and the x axis represents the value of the systemic property in the reference model. Techniques involving moving quantiles are introduced to generate secondary plots in which correlations and inhomogeneities in correlations are more easily detected. Several examples that illustrate the advantages of these techniques are presented and discussed.*

Contact: *Savageau@umich.edu*

Introduction

The only rigorous way to characterize and compare alternative biological designs for a particular class of systems is through the use of mathematical models and quantitative methods of analysis. In pursuing these goals we must address three critical issues. First, biologically meaningful behaviors must be identifed (or, as is more commonly the case, hypothesized) and characterized by quantitative measures. Second, a representation of the alternatives must be capable of describing the phenomena of interest in quantitative terms. Third, comparisons will require analyses that explore a range of parameter values and use statistical methods to evaluate the results.

The first issue is obviously critical if the results are to be biologically significant; however, there is no prescription for discovering those biological behaviors that are based on natural selection or those that occur at random with high probability. The behaviors that are important characteristics of a given biological system can only be discovered by experimental means. Hypotheses must be generated and tested in each case, and this process will vary considerably according to the systems being studied. The behavioral repertoire of nonlinear systems can be quite diverse including saturation, thresholds, memory, time delays, synchrony, stable limit cycles and strange attractors.

The second issue is critical to any quantitative comparison of alternative systems. We require a mathematical language (or formalism) that is sufficiently flexible to represent the diverse behaviors that are likely to be encountered in the quantitative description of a nonlinear biological system. The power-law formalism (Savageau, 1996) is a most likely candidate for this language. It can be viewed as a canonical nonlinear representation from three different perspectives. From a fundamental perspective, it provides a generalization of mass-action kinetics, which is the most widely used representation of biological systems at the molecular level. From a recasting perspective, it provides a globally accurate representation that can be made mathematically equivalent to any sufficiently differentiable nonlinear system. From a local perspective, it provides a general representation that is guaranteed to be accurate over a range of variation about a nominal operating point.

*To whom correspondence should be addressed.

© Oxford University Press 2000

The third issue is critical because values for many of the parameters in any given complex system will not have been measured, and for those that have the estimates will often be poor. Moreover, even if we had a complete set of accurate parameter values with which to study the behavior of a system, the results would only apply to that particular system. In any case, we would have to vary the parameters over a range of values and statistically analyze the results to determine the properties of the general class of systems to which the particular system belongs.

Our purpose in this and the following paper is to present a methodology for dealing with this third issue and to illustrate its use in the simplest setting where the essentials of the methodology can be made most transparent. Hence, we shall focus on a class of systems for which the biologically relevant behavior is relatively simple and well defined (namely, unbranched amino-acid biosynthetic pathways with a single homeostatic steady state) and for which the local nonlinear representation, which is the simplest of the representations within the power-law formalism, is appropriate. At the end of the second paper we will return to these issues and indicate how the methods presented here might be applied to systems with more complex behaviors requiring more general representations within the power-law formalism. The methods themselves provide an extension of a previously developed approach for making well-controlled comparisons.

In the study of complex biological networks, models with alternative designs or structure are often compared to determine which of them provides the better representation for some observed phenomenon (e.g. Ni and Savageau, 1996). When comparing structurally different models for the same phenomenon, it is difficult to know whether the differences observed are accidental or inherent differences that can be attributed specifically to the alternative designs. The method of Mathematically Controlled Comparison (Savageau, 1972; for a review see Irvine, 1991) was proposed to address this issue.

In brief, the steps involved in this method are as follows. First, mathematical models are formulated for the alternative designs being compared. For example, a biosynthetic pathway with end-product inhibition and an identical one without it. One model, generally the more complex, is designated the *reference*; the other is designated the *alternative*. Second, the parameters of the alternative model are fixed relative to those of the reference model. Each process in the alternative model that is identical to one in the reference model is assigned a set of parameter values that is identical to the corresponding set in the reference model. This is referred to as *internal equivalence*. Each process in the alternative model that is different from the corresponding process in the reference model will have a set of parameter values that is unique to the alternative model, and these parameters represent degrees of freedom that must be constrained in an effort to reduce the accidental differences between the models. Each constraint is established by equating the expressions for a systemic property common to the two models. The set of constraint equations is then solved to determine values for the unique parameters of the alternative model in terms of values for the parameters of the reference model. This is referred to as *external equivalence*. Finally, having eliminated all the degrees of freedom, the two models are analyzed to determine the differences that remain.

The critical step in this method is the solution of the constraint equations. The models are described by nonlinear equations that in general have no analytical solution. However, the discovery of a canonical nonlinear representation that is locally valid and amenable to analytical solution (Savageau, 1969a, 1969b; for a review see Savageau, 1996) removes the difficulty associated with this critical step in many cases (Savageau, 1972, 1976). This canonical nonlinear representation within the power-law formalism is referred to as an S-system and it has the following systematic structure:

$$\frac{dX_i}{dt} = \alpha_i \prod_{j=1}^{n} X_j^{g_{ij}} - \beta_i \prod_{j=1}^{n} X_j^{h_{ij}} \qquad i = 1, 2, \ldots, n. \tag{1}$$

For each dependent concentration X_i in a biochemical model there exists an aggregate production function and an aggregate consumption function. These aggregate functions are approximated by a first-order Taylor series in a logarithmic space, which in Cartesian space leads to the product of power-law functions. An exponent of zero for any X_j means that that variable has no direct influence on the rate of the corresponding aggregate process, a positive exponent means that the variable and the rate of the aggregate process are positively correlated, and a negative exponent means that they are negatively correlated. In a steady state, (1) becomes a linear equation in logarithmic space and can be solved analytically. Likewise, various systemic properties can be calculated analytically and used to form constraints by equating the analytical expressions for corresponding systemic properties in the two models. These constraint equations can then be solved to determine values for the unique parameters of the alternative model in terms of values for the parameters of the reference model.

Once internal and external equivalence between the models is established in this manner, we can proceed to analyze the models and compare their systemic behaviors by taking ratios of their corresponding properties. The steady-state properties that are typically analyzed in Mathematically Controlled Comparisons include concentrations, fluxes, logarithmic gains, parameter sensitivities, and stability margins. For the purposes of this paper, these

systemic properties will be represented by M. The ratio of M in the reference model to M in the alternative model exhibits one of three possible properties.

1. The analytical ratio of M values is always equal to 1, independent of parameter values. This means that the property being analyzed is always the same in the two models.
2. The analytical ratio of M values is always larger (smaller) than 1, independent of parameter values. This means that the property being analyzed is always larger (smaller) in the reference model than in the alternative model. However, if the numerical values for the parameters are not known we can not say how much larger (smaller) the property is.
3. The analytical ratio of M values is larger or smaller than 1, depending on the parameter values. In this case it is difficult to say anything about the property by simple examination of the analytical ratio.

The uncertainties associated with properties 2 and 3 will be addressed by the numerical methods being proposed in this paper. Moreover, these methods will allow us to draw statistical conclusions about the relative merits of various biological designs.

Methods

If we knew the numerical values for all the parameters of the reference model, then we could calculate the numerical values for all the parameters of the alternative model that is internally and externally equivalent. However, knowledge of all the parameter values is rarely available for any model. Furthermore, using just one set of parameter values restricts the interpretation to the specific pair of models being compared. These limitations can be overcome by creating a large ensemble of reference models with randomly generated sets of parameter values that adequately sample the parameter space. For each of these one can then construct the alternative model that is internally and externally equivalent.

There are two types of parameters that appear in the S-system representation (equation (1)): exponential parameters (kinetic orders) and multiplicative parameters (rate constants). The exponential parameters, which are weighted averages of more elementary kinetic orders, typically have values less than 4 in magnitude (Voit and Savageau, 1987). The multiplicative parameters, which reflect the different time scales present within the model, for most cases of interest are within 4 orders of magnitude of each other (i.e. within 4 $\log_{10}$ units). The results given in the following section are not critically dependent upon this particular choice of limits for the parameter space that needs to be sampled.

By using randomly generated numbers we can sample the relevant parameter space, apply selection and create a large ensemble of biologically relevant numerical models for both the reference and alternative designs, and make an ensemble of numerical comparisons. The amount of data generated by this approach can be overwhelming. The following subsections describe several ways to treat and interpret these data. In a following paper (Alves and Savageau, 2000) these methods are applied to a specific class of biochemical control mechanisms in a context different from that of mathematically controlled comparisons. Subsequent papers will provide examples of specific applications within the mathematically controlled comparison framework.

Basic treatment and analysis of the comparisons

The first problem in analyzing a large number of comparisons is deciding how to represent the data. Since we are comparing the value of a given property M between the reference model and its alternative, one obvious way to represent the data is by taking the ratio of M in the reference model to M in the alternative model.

$$R = M_{\text{reference}}/M_{\text{alternative}}. \tag{2}$$

When dealing with an ensemble of comparisons we must calculate the ratio, R, of M values for each reference model and its alternative model that is internally and externally equivalent. These data then can be treated by calculating some quantile of interest for the ensemble of ratios, thus determining whether M is statistically larger in the reference models or their alternatives. This, however, will not give us much information, even if we included calculations for the dispersion of the results.

Density plots

More information can be obtained from density plots of R versus M, where M is a property measured in the reference model; e.g. the sensitivity, $S(X_i, \alpha_1)$, of a given intermediate, X_i, to fluctuations in the rate constant, α_1, for the first reaction of the pathway. Some density plots where the ratio is typically smaller than 1 are presented in Figures 1–3. Note that in Figure 1A we have a situation in which the ratio of $S(X_i, \alpha_1)$ is uniformly scattered throughout the entire region bounded by $R = 1$ and $R = 0$. Figures 2A and 3A show different non-traditional distributions. Figure 3A shows a case in which M can take only discrete values.

Density plots can be used to determine rank correlations between M and R. Traditionally we calculate non-parametric rank correlations by using point measures such as the Spearman or Kendal rank correlation coefficients (e.g. Wherry, 1984; Krauth, 1988). These methods find linear and non-linear rank correlations between variables; however, it is not always easy to find such correlations in

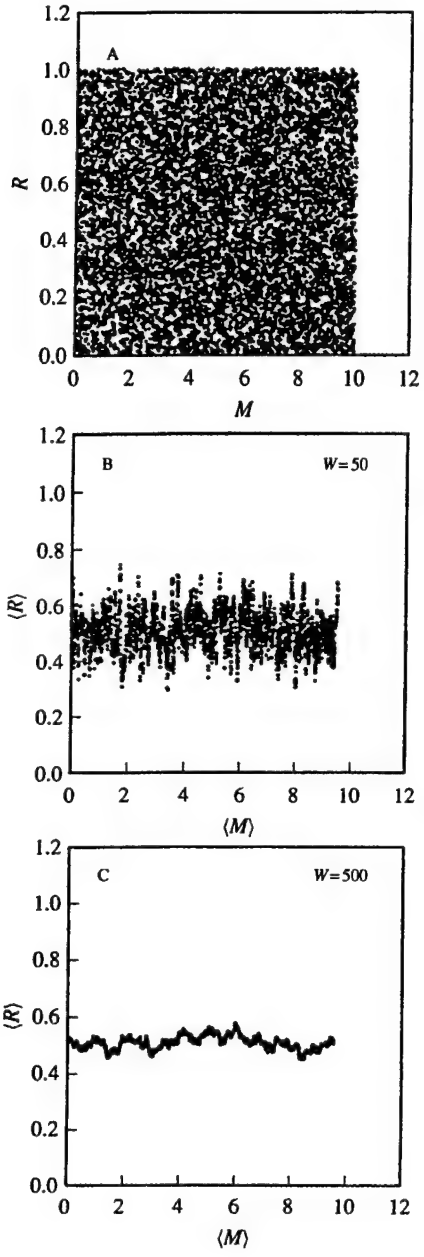


Fig. 1. Uncorrelated Density of Ratios Plots. A: Density Plot of R versus M for two alternative models. There is a uniform distribution of values on both axies. B: Moving median plot of $\langle R \rangle$ versus $\langle M \rangle$ for the data in panel A and a window size of $W = 50$. C: Moving median plot of $\langle R \rangle$ versus $\langle M \rangle$ for the data in panel A and a window size of $W = 500$. See text for discussion.

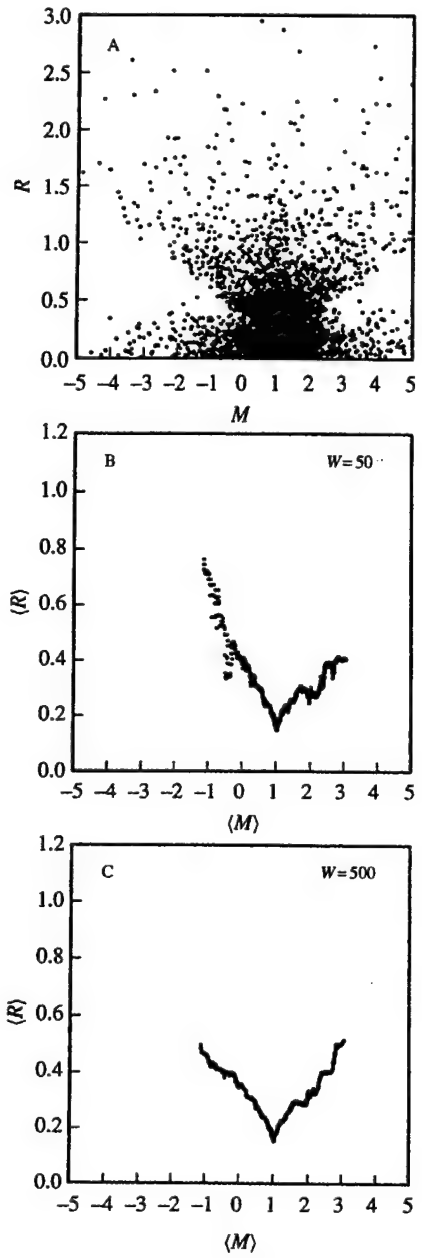


Fig. 2. Correlated Density of Ratios Plots. A: Density Plot of R versus M for two alternative models. B: Moving median plot of $\langle R \rangle$ versus $\langle M \rangle$ for the data in panel A and a window size of $W = 50$. C: Moving median plot of $\langle R \rangle$ versus $\langle M \rangle$ for the data in panel A and a window size of $W = 500$. See text for discussion.

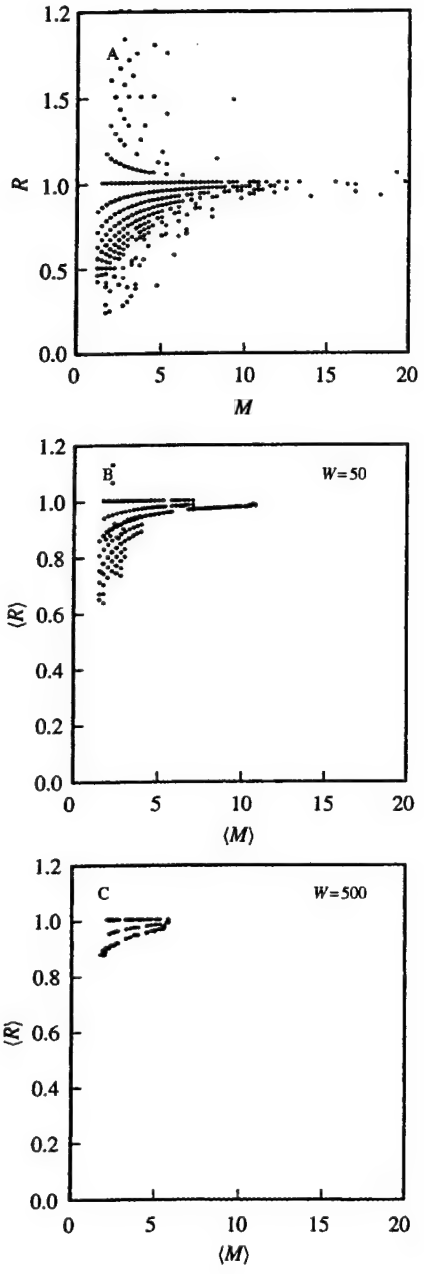


Fig. 3. Correlated Discrete Density of Ratios Plots. A: Density Plot of R versus M for two alternative models for which M and R assume discrete values. B: Moving median plot of $\langle R \rangle$ versus $\langle M \rangle$ for the data in panel A and a window size of $W = 50$. C: Moving median plot of $\langle R \rangle$ versus $\langle M \rangle$ for the data in panel A and a window size of $W = 500$. See text for discussion.

density plots that are as scattered as the ones presented in Figures 1A or 2A.

The analysis of these density plots using point measures can be almost as cumbersome to interpret as the results from purely symbolic analysis. Furthermore, the point measures will almost certainly hide information that would be available from a less coarse analysis. The frequency of different values in a density plot is typically analyzed using two- and three-dimensional histograms. However, this approach may or may not lead to the determination of standard statistical distributions that fit the pattern of the data.

Quantile analysis of density plots

Moving quantile techniques allow us to interpret density plots using either parametric or non-parametric statistics. However, we shall refer only to the non-parametric case in the remainder of this paper. Let us assume that we want to know whether the M values of the reference model are larger than those of the alternative model (i.e. $R > 1$) more often than not. This can be determined from the median of the ratios, i.e. Quantile 0.5 ($Q_{0.5}$), which will be denoted $\langle R \rangle$. If for some reason we want to know whether R is greater than 1 in more than 80% of the cases, then we would be dealing with $Q_{0.8}$. For the rest of this paper we will consider only plots of $\langle R \rangle$ for reasons of simplicity. The correlation between magnitude M and ratio R can be obtained from the moving quantile technique instead of from the point measures technique mentioned above.

The density plot can be viewed as a list of N paired values. Initially we order the pairs with respect to the reference magnitude to form a list L_1 in which the first pair has the lowest measured value for M in the reference model, the second has the second lowest and so on. Next we build a secondary plot as follows.

Pick a window size W smaller (usually much smaller) than the sample size N, collect the first W ratios from the list L_1, calculate the $Q_{0.5}$, and pair this number $\langle R \rangle$ with the median value of the corresponding M values of the reference model, which will be denoted $\langle M \rangle$. Advance the window by one position, collect ratios 2 to $W + 1$, calculate $\langle R \rangle$, and pair it with the corresponding $\langle M \rangle$ value. Continue this procedure until the last ratio from the list L_1 is used for the first time. We now have a new list, L_2, of size $N - W + 1$ that is ordered from the smallest to largest values for $\langle M \rangle$ of the reference model. A moving median exhibits the following general statistical properties. For an infinite ordered population, the moving median tends to the mean of the population as the window size W increases without limit. For a finite ordered sample of size N, the moving median tends to the median of the sample as W approaches N.

The plot of list L_2 exhibits a moving median $\langle R \rangle$ on the y axis that corresponds to the equivalent moving median

$\langle M \rangle$ on the x axis. These moving quantile plots allow us to determine the percentage of comparisons in which $\langle R \rangle$ is larger than 1 and, at the same time, whether or not there is any correlation between $\langle R \rangle$ and $\langle M \rangle$ of the reference model. A slope of zero or infinity in the moving quantile plot of L_2 shows there is no correlation between $\langle R \rangle$ and $\langle M \rangle$ of the reference model. Applications of moving average techniques that are of a more classical nature can be found in Hamilton (1994) and Huang and Dunsmuir (1998).

Examples and discussion

Moving median plots of L_2 lists can be used to compare the relative effectiveness of two different classes of models on the basis of some criterion. For example, assume that M measures the sensitivity of a model to fluctuations in a given parameter and that this parameter sensitivity should be as low as possible according to the criterion of model robustness. The ratio R of M values in the reference model to M values in the alternative model, which is otherwise internally and externally equivalent to the reference model, is plotted and from this density plot one forms the moving median plot of $\langle R \rangle$ versus $\langle M \rangle$. Examples of such plots that exhibit various patterns are presented and their interpretation discussed below.

Figure 1A shows a plot of R versus M with $N = 10000$ and a uniform scatter in both the R and M values. Since the scatter is uniform, we would expect to find that $\langle R \rangle$ is independent of $\langle M \rangle$. In this example M values are in the interval [0, 10] and R values are in the interval [0, 1]. As the window size grows, the resulting moving median $\langle R \rangle$ approaches 0.5 with progressively smaller bounds because 0.5 is the median of the sample. Figures 1B and C show plots of $\langle R \rangle$ versus $\langle M \rangle$ for window sizes of 50 and 500, which can be thought of as the relevant sample size in this context. Figures 1B and C also show that there is no correlation between $\langle R \rangle$ and $\langle M \rangle$; i.e. the values for the moving median $\langle R \rangle$ are independent of the values for $\langle M \rangle$ of the reference model.

It is important to emphasize that different density plots can have similar moving quantile plots, due to the statistical nature of quantiles. For example, if R and M were both normally distributed and uncorrelated, then the moving median plots of $\langle R \rangle$ versus $\langle M \rangle$ would be similar to those in Figures 1B and C for the same sample and window sizes.

Figure 2A shows a plot in which R is sometimes larger than 1. However, the moving quantile plot for $Q_{0.5}$ in Figures 2B and C show that in most cases the value for M of the reference model is smaller than that of the alternative model ($\langle R \rangle$ less than 1). Also, there is a clear correlation between the value for $\langle R \rangle$ and the value for $\langle M \rangle$ of the reference model, which is unlike the case in Figures 1B and C. The value for $\langle R \rangle$ is a function of $\langle M \rangle$ with a minimum around $\langle M \rangle \approx 1$. With $\langle M \rangle \approx 1$, the value for M of the reference model is much less than that of the alternative model. With values for $\langle M \rangle$ that are increasing or decreasing away from 1, the value for M of the reference model approaches that of the alternative model.

The selection of an appropriate window size is critical. If W is too small (e.g. 5), the $Q_{0.5}$ plot will not differ significantly from the raw density plot. If W is too large, the correlation between $\langle R \rangle$ and $\langle M \rangle$ will be lost, or at least attenuated. This can be seen by comparing the curves for the two different window sizes in Figures 2B and C. As the window size increases from 50 to 500, the slope of the branch for $\langle M \rangle$ less than 1 decreases (if the window size is increased further, the slope eventually becomes 0). This happens because the early samples of R are contaminated with latter samples and the correlation with the lower values of M is lost. With larger window sizes the slope of the branch for $\langle M \rangle$ greater than 1 also decreases. As W approaches N, the slope of the curve on either side of $\langle M \rangle \approx 1$ tends toward 0 and the $Q_{0.5}$ plot provides no more information than calculating the median of the entire sample. Thus, the advantages of a $Q_{0.5}$ plot only become apparent at intermediate window sizes. There is, to our knowledge, no good way of deciding the optimal size for the window W; this depends on the sample size N and on the nature of the sample itself and must be determined by trial and error.

Figure 3A illustrates a case in which the values of M can only assume a finite number of discrete values. Figures 3B and C show the corresponding plots for $\langle R \rangle$ versus $\langle M \rangle$ of the reference model. A correlation between $\langle R \rangle$ and $\langle M \rangle$ is evident at low values of $\langle M \rangle$ but disappears as $\langle M \rangle$ increases. In addition, the $Q_{0.5}$ plot in Figures 3B and C shows the dispersion in the moving median at each value of $\langle M \rangle$, unlike the $Q_{0.5}$ plots in Figures 1B and C and in Figures 2B and C. This dispersion occurs because there are several pairs in the list L_1 that have the same discrete value for M but different discrete values for R. As the window W moves through a series of identical M values, the median value for M will remain unchanged whereas the median value for R will change. One can construct discrete density plots for any of the previous examples by designating classes for the values of M and by representing each class by the median of the class interval. Thus, the plot of $\langle R \rangle$ versus $\langle M \rangle$ in cases such as these can give us information not only about frequencies and correlations but also about dispersion of the results.

Acknowledgments

This work was supported in part by a joint PhD fellowship PRAXIS XXI/BD/9803/96 granted by PRAXIS XXI

through Programa Gulbenkian de Doutoramentos em Biologia e Medicina (RA), the US Public Health Service Grant RO1-GM30054 from the National Institutes of Health (MAS), and the US Department of Defense Grant N00014-97-1-0364 from the Office of Naval Research (MAS). We thank Armindo Salvador, Rui Gardner Oliveira, and Charlotte Eriksson for critically reading early versions of this manuscript and making useful comments.

References

Alves,R. and Savageau,M.A. (2000) Systemic properties of ensembles of metabolic networks: application of graphical and statistical methods to simple unbranched pathways. *Bioinformatics*, **16**, 534–547.

Hamilton,J.D. (1994) *Time Series Analysis*. 2nd edn, Princeton University Press, New Jersey.

Huang,D. and Dunsmuir,W.T.M. (1998) Computing joint distributions of 2D moving median filters with applications to detection of edges. *IEEE Trans. Patt. Anal. Mach. Intel.*, **20**, 340–343.

Irvine,D.H. (1991) The method of controlled mathematical comparisons. In Voit,E.O. (ed.), *Canonical Nonlinear Modeling: S-Systems Approach to Understanding Complexity*. Van Nostrand Reinhold, New York, pp. 90–109.

Krauth,J. (1988) *Distribution-Free Statistics: An Application-Oriented Approach*. Elsevier, New York.

Ni,T.-C. and Savageau,M.A. (1996) Model assessment and refinement using strategies from biochemical systems theory: application to metabolism in human red blood cells. *J. Theoret. Biol.*, **179**, 329–368.

Savageau,M.A. (1969a) Biochemical systems analysis I: some mathematical properties of the rate law for the component enzymatic reactions. *J. Theoret. Biol.*, **25**, 365–369.

Savageau,M.A. (1969b) Biochemical systems analysis II: the steady state solution for an n-pool system using a power law approximation. *J. Theoret. Biol.*, **25**, 370–379.

Savageau,M.A. (1972) The behavior of intact biochemical control systems. *Curr. Top. Cell Reg.*, **6**, 63–130.

Savageau,M.A. (1976) *Biochemical Systems Analysis: A Study of Function and Design in Molecular Biology*. Addison-Wesley, Reading, Mass.

Savageau,M.A. (1996) Power-law formalism: a canonical nonlinear approach to modeling and analysis. In Lakshmikantham,V. (ed.), *World Congress of Nonlinear Analysts 92, Vol. 4*. Walter de Gruyter Publishers, Berlin, pp. 3323–3334.

Voit,E.O. and Savageau,M.A. (1987) Accuracy of alternative representations for integrated biochemical systems. *Biochemistry*, **26**, 6869–6880.

Wherry,R.J. (1984) *Contributions to Correlational Analysis*. Academic Press, Orlando.

BIOINFORMATICS

Vol. 16 no. 6 2000
Pages 534–547

Systemic properties of ensembles of metabolic networks: application of graphical and statistical methods to simple unbranched pathways

Rui Alves [1, 2, 3] and Michael A. Savageau [1,*]

[1]Department of Microbiology and Immunology, University of Michigan Medical School, 5641 Medical Science Building II Ann Arbor, MI 48109-0620 USA, [2]Grupo de Bioquimica e Biologia Teoricas, Instituto Rocha Cabral, Calçada Bento da Rocha Cabral 14, 1250 Lisboa, Portugal and [3]Programa Gulbenkian de Doutoramentos em Biologia e Medicina, Departamento de Ensino, Instituto Gulbenkian de Ciência, Rua da Quinta Grande 6, 1800 Oeiras, Portugal

Received on July 28, 1999; revised on November 11, 1999; accepted on December 3, 1999

Abstract

Motivation: *Mathematical models are the only realistic method for representing the integrated dynamic behavior of complex biochemical networks. However, it is difficult to obtain a consistent set of values for the parameters that characterize such a model. Even when a set of parameter values exists, the accuracy of the individual values is questionable. Therefore, we were motivated to explore statistical techniques for analyzing the properties of a given model when knowledge of the actual parameter values is lacking.*

Results: *The graphical and statistical methods presented in the previous paper are applied here to simple unbranched biosynthetic pathways subject to control by feedback inhibition. We represent these pathways within a canonical nonlinear formalism that provides a regular structure that is convenient for randomly sampling the parameter space. After constructing a large ensemble of randomly generated sets of parameter values, the structural and behavioral properties of the model with these parameter sets are examined statistically and classified. The results of our analysis demonstrate that certain properties of these systems are strongly correlated, thereby revealing aspects of organization that are highly probable independent of selection. Finally, we show how specification of a given behavior affects the distribution of acceptable parameter values.*

Contact: *Savageau@umich.edu*

Introduction

The characterization of large and complex biochemical networks cannot be achieved with the direct intuitive approaches that have been successful for simpler model systems. The more systematic tools provided by mathematical modeling and computer analysis have become essential because they are especially well suited for organizing large amounts of data and representing nonlinear and parallel processes.

The most common method of constructing an appropriate model for a biochemical system has been the reductionist or bottom-up approach. The component parts are isolated and characterized, and then the resulting submodels are assembled into a model of the integrated system. For example, in the study of metabolic pathways, individual enzymes were isolated and kinetically characterized *in vitro*; pathway models were then constructed by assembly of the individual rate laws. The fundamental problems inherent in this approach are three (Ni and Savageau, 1996):

1. failure to identify all the relevant components
2. failure to identify all the relevant interactions
3. failure to determine accurately all the relevant parameter values.

The associated practical problems are the enormous numbers of components and interactions that need to be identified and the difficulty of reproducing the conditions experienced by the components in their natural setting so that their parameter values can be accurately determined *in vitro* (e.g. Clegg, 1984; Moore *et al.*, 1984; Ovadi and Srere, 1996; Savageau, 1992; Sorribas *et al.*, 1993). These problems have limited the success of the bottom-up approach (e.g. Albe and Wright, 1992; Antunes *et al.*, 1996; Curto *et al.*, 1998; Ni and Savageau, 1996; Shiraishi and Savageau, 1993)

An alternative method of constructing an appropriate model is often termed the reverse-engineering or top-down approach. Many of the variables are measured in

*To whom correspondence should be addressed.

© Oxford University Press 2000

the intact system, and then one attempts to reconstruct the underlying model that produced these data. The fundamental problems in this case are:

1. selecting a mathematical representation that is sufficiently general so that one can be assured that it will encompass the system to be characterized
2. the theoretical limits on what can be identified (problem of identifiability: e.g. Chappell and Godfrey, 1992; Feng and DiStefano, 1991; Ginn and Cushman, 1992) when one can only measure a subset of the variables (problem of observability: e.g. Moheimani *et al.*, 1996; Xu *et al.*, 1996).

The practical problems are associated with the limits of the technologies currently available for measuring all the relevant variables. Although the top-down approach has long been applied in simple cases that illustrate the method (e.g. Brown *et al.*, 1990; Diamond, 1975; Domnitz, 1976; Kargi and Shuler, 1979; Quant, 1993; Voit and Savageau, 1982), its use in biology is currently being driven by the new techniques coming out of the Human Genome Project that generate massive data sets (e.g. Brown and Botstein, 1999; Chu *et al.*, 1998; DeRisi *et al.*, 1997; Eisen *et al.*, 1998; Somogyi *et al.*, 1997; Törönen *et al.*, 1999). Although the top-down approach shows considerable promise, it is unlikely that this method alone will provide a satisfactory solution to the problem of modeling large and complex biochemical systems.

If one is interested in modeling a specific system (e.g. tryptophane biosynthetic pathway of *Escherichia coli*), the best way to proceed is to measure all the necessary parameters of the system in the organism of interest and build the model based on those values. One could productively combine the bottom-up and top-down approaches described above (Bliss *et al.*, 1982; Yanofsky and Horn, 1994). On the other hand, if one is interested in a generic class of systems (e.g. amino acid biosynthetic pathways in general) or if the measurements are impossible to perform with accuracy and precision, even a combination of the two approaches may not be adequate.

In this paper we propose a statistical approach for dealing with generic classes of biochemical systems. We apply this approach to a general three-step unbranched biosynthetic pathway with inhibitory feedback. This pathway is an abstraction from the collection of unbranched pathways responsible for the biosynthesis of amino acids (e.g. see http://www.genome.ad.jp/kegg/dblinks/map/map01150.html). The results of our analysis demonstrate that certain properties of these systems are strongly correlated, thereby revealing aspects of organization that are highly probable independent of selection.

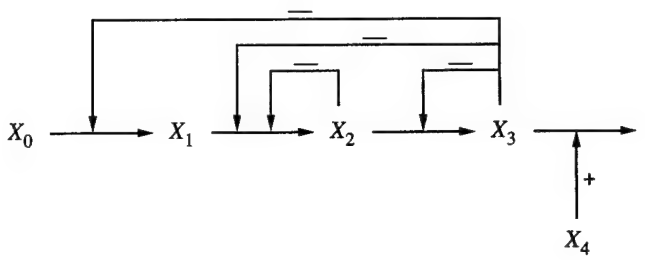


Fig. 1. Three-step unbranched biosynthetic pathway with inhibitory feedback. The metabolites are represented by X with an appropriate subscript. The horizontal arrows represent chemical conversion, whereas the vertical arrows represent modifier influences either positive or negative. This pathway can be viewed as an abstraction of the biosynthetic pathways for amino acids.

Methods

Amino acid biosynthetic pathways and their regulation have been studied intensively for more than 40 years. There is widespread acceptance among cell physiologists that the principal role of these systems is to provide a homeostatically regulated supply of amino acid for protein synthesis. This role has been characterized in terms of several behaviors that can be described by quantitative criteria (Savageau, 1976) that will be elaborated upon in this paper.

Systemic description and analysis

An unbranched three-step pathway with feedback inhibition is depicted in Figure 1. The independent variable X_4 represents the cell demand for the end product X_3. If the cell requires large amounts of X_3, then the value of X_4 will be high; if small amounts of X_3 are required, then the value of X_4 will be low. The dynamic behavior of such a model can be described by a set of ordinary differential equations, one equation per intermediate. This set of equations can be approximated to the first order in logarithmic space, yielding another set of ordinary differential equations with the canonical form of an S-system (Savageau, 1969):

$$\begin{aligned}
\frac{dX_1}{dt} &= \alpha_1 X_0^{g_{10}} X_3^{g_{13}} - \alpha_2 \prod_{j=1}^{3} X_j^{g_{2j}} \\
\frac{dX_2}{dt} &= \alpha_2 \prod_{j=1}^{3} X_j^{g_{2j}} - \alpha_3 \prod_{j=2}^{3} X_j^{g_{3j}} \\
\frac{dX_3}{dt} &= \alpha_3 \prod_{j=2}^{3} X_j^{g_{3j}} - \alpha_4 X_3^{g_{43}} X_4^{g_{44}}.
\end{aligned} \tag{1}$$

The multiplicative parameters, α, can be interpreted as rate constants that are always positive. The exponential parameters, g, can be interpreted as kinetic orders that

represent the direct influence of each species on each rate law. If X_i is directly involved in the reactions of the aggregate rate law V_j, as either a substrate or a modulator, and if an increase in X_i causes an increase in the rate V_j, then the kinetic order will be positive. If an increase in X_i causes a decrease in V_j, then the kinetic order will be negative. If X_i is not directly involved in V_j, then the kinetic order will be zero. The kinetic orders $g_{i+1,i} (0 \leq i \leq 3)$ in (1) are positive because these are the kinetic orders for substrates of reactions. The kinetic order g_{44} will be set arbitrarily equal to 1 for the remainder of this paper in order to simplify subsequent calculations. This will not affect our results in any significant way since g_{44} is simply a scale factor in logarithmic space. The remaining kinetic orders, which represent negative feedback interactions, are negative.

At a steady state, the rate of production and the rate of consumption will be equal for each intermediate, and (1) reduces to the following matrix equation (Savageau, 1969):

$$\begin{bmatrix} b_1 - g_{10}Y_0 \\ b_2 \\ b_3 + Y_4 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} \quad (2)$$

where $b_i = \ln(\alpha_{i+1}/\alpha_i)$, $a_{ij} = g_{ij} - g_{i+1,j}$ and $Y_i = \ln(X_i)$. This linear equation is easily solved, e.g. using Cramer's rule, to provide a steady-state expression in symbolic form for each Y_i.

Other steady-state magnitudes of interest can be calculated in a similar way. Logarithmic gains quantify the influence of each independent variable on each dependent variable; e.g. the logarithmic gain

$$L(X_i, X_0) = \frac{\mathrm{d}\ln(X_i)}{\mathrm{d}\ln(X_0)} = \frac{\mathrm{d}Y_i}{\mathrm{d}Y_0}$$

gives the percentage change in an intermediate X_i caused by a percentage change in X_0. These logarithmic gains are calculated analytically at the steady state (Savageau, 1971) by differentiating each Y_i with respect to Y_0.

Parameter sensitivities quantify the influence of each parameter on each dependent variable of the system; e.g. the sensitivity

$$S(X_i, p_j) = \frac{\mathrm{d}\ln(X_i)}{\mathrm{d}\ln(p_j)} = p_j \frac{\mathrm{d}Y_i}{\mathrm{d}p_j}$$

gives the percentage change in the concentration X_i caused by a percentage change in the parameter p_j. These parameter sensitivities and those of the steady-state flux are also calculated analytically at the steady state. The parameter sensitivities give important information about the sensitivity of the system to perturbations in its structure.

The steady state for an unbranched biosynthetic pathway should be locally stable; i.e. the system should return to its original steady state after a small perturbation in the variables (as opposed to the parameters) of the system. If this does not occur, the system is dysfunctional. The stability can be determined by using the well-known Routh criteria (Savageau, 1976).

Any of these systemic properties can be analytically determined in the steady state by using the S-systems local representation. However, having an analytical expression for these systemic properties is just the first step in the analysis of a system. Interpretation of these analytical expressions can be problematic because they depend on many parameters and their behavior is too complex for easy visualization. Even when a general qualitative interpretation can be obtained just by looking at the closed-form expressions [e.g. $L(X_3, X_0) < L(X_1, X_0)$], the results are difficult to quantify [e.g. how much larger is $L(X_1, X_0)$?].

Also, there are no general closed-form solutions for the dynamic properties of the system. To analyze these properties one must specify numerical values for the parameters and solve the differential equations (1) using numerical techniques. An example of such a property is the settling time of a system, which is defined as the time required for a system to return to its steady state after a perturbation in the levels of its metabolites. The settling time also gives us an indication of the average transit time for material passing through the system. Short transit times allow a system to respond rapidly to changes in its environment (Savageau, 1972).

Defining classes of systems for statistical comparison

If one wishes to understand the general properties of pathways such as the one depicted in Figure 1, then one faces the following dilemma. General results that follow from the closed-form analytical expressions may be too complex to interpret and quantify, and quantitative results for particular values of the parameters do not yield general insights. One way of resolving this dilemma is to study the statistical properties for a class of systems generated by an ensemble of sets of parameter values. We shall consider two different methods for defining the class of interest.

Structural classes. Systems that have the same network topology (i.e. have the same pattern of interactions among their elements and the same signs for the interactions) will be defined as members of the same structural class. As a case study for this paper we have chosen the system in Figure 1 and described its local behavior by the S-system representation within the power-law formalism. By so doing we have defined a specific class of systems that share the same network topology. By focusing on such a topology we have limited the study to systems belonging

to the same structural class. Individual members of this structural class can be generated by sampling the space of parameters that define the class and their characteristics can be obtained from the corresponding solutions of (1).

Behavioral classes. Systems that exhibit a specific type of systemic behavior will be defined as members of the same behavioral class. For example, those systems belonging to the structural class in Figure 1 that have a single locally stable steady state can be defined as members of a behavioral class. Individual members of such a behavioral class cannot be generated directly by sampling at random the space of parameters because some of the parameter sets will produce unstable systems. Instead, they must be generated indirectly, e.g. by sampling at random the space of parameters, testing the sample for the desired behavior, and then retaining only the relevant samples.

In the example above the behavioral class is a subclass within the structural class, but this need not be so. If our only knowledge of the system was that it had three metabolites, we could study an ensemble of models in which each kinetic order might have positive or negative values, which generates models belonging to different structural classes. One could then choose models for study based simply on their behavior, disregarding the signs of the kinetic orders.

Several (elementary) behavioral classes can be combined to define a composite behavioral class whose members are systems that exhibit all of the individual systemic behaviors.

Sampling the parameter space

The regular structure of the local S-system representation facilitates building the ensemble of sets of parameter values. The positive kinetic orders $g_{i+1,i}$ refer to enzymes binding their substrates. The maximum value for these kinetic orders is given by the number of substrate binding sites on the enzyme.[†] In the majority of cases there are less than four such sites (Hlavacek and Savageau, 1995; Voit and Savageau, 1987). Thus we will assume that these kinetic orders have values between 0 and 5. The negative kinetic orders (g_{13}, g_{22}, g_{23} and g_{33}) refer to enzymes binding inhibitors. In most cases there are again fewer than four such binding sites per enzyme, and we will assume that these kinetic orders have values between -5 and 0. One can always normalize the time scale with respect to one of the rate constants. The others will be assumed to have normalized values within 5 orders of magnitude of 1. Thus, the logarithm of each normalized rate constant will have values between -5 and 5.

In building an appropriate ensemble of sets of parameter values one needs to use a representative sample of the allowable parameter space. Since the statistical distribution of parameter values in real-life systems is unknown, the most appropriate approach is to sample the space uniformly. There are several strategies for accomplishing this.

First, one can impose a regular grid on the multidimensional parameter space and use the vertices of that grid to define the set of parameter values. In general, a system with n unknown parameters and the same grid size, ω, will require ω^n samples. This exponential increase in number of required samples makes it difficult to maintain a dense grid as the number of parameters increases. Also, maintaining a rigid grid complicates matters when one is studying ensembles of parameter sets that give rise to certain types of systemic behavior. Second, pseudo-random number generators can be used to generate the largest possible sample size without having a rigid grid to sample from. This method facilitates the study of ensembles of parameter sets that give rise to certain types of systemic behavior. Third, strategies based on number theory can be used to generate what are known as quasi-random numbers that are uniformly distributed. Examples include Halton and Solov sequences [for a review see Bratley and Fox (1988)]. Finally, another technique devised for dealing with large parameter spaces is the Latin Hyper cube. The Latin Hyper-cube ensures that each parameter will be sampled in every one of its sub-ranges. It has no advantage over the other methods mentioned above if there are important interactions between parameters [for a discussion see Dunn and Clark (1974)]. For the results reported below we have used the pseudo-random number generator.

Specifying behavioral classes

Since the system in Figure 1 is an abstraction of an unbranched biosynthetic pathway, the literature was searched and a basic number of desirable characteristics have been found for such systems. The group of all these characteristics was used to define a composite behavioral class. If the model generated by a given set of parameter values did not belong to this class, then the set was discarded and a new random set was tested. In this way we generated ensembles of 5000 for our studies.

The composite behavioral class studied is defined by a collection of six elementary behavioral classes with the following characteristics:

B1. The steady-state concentration of pathway intermediates should be low when compared with the concentration of the final product. The major function of unbranched biosynthetic pathways is production of their end product (e.g. X_3 in the example of Figure 1). High concentrations of intermediates *per se*

[†] This is not always true of reversible reactions operating close to equilibrium. The usual strategy for aggregating fluxes can lead to kinetic orders with extremely large absolute values. This problem can be solved by using an alternative strategy for aggregating fluxes (Sorribas and Savageau, 1989). However, we will not deal with these cases here.

are unnecessary; they would tax the solvent capacity of the cell and potentially interfere in a nonspecific way with otherwise unrelated reactions (e.g. Atkinson, 1969; Savageau, 1972; Srere, 1987 and Levine and Ginsburg, 1985, for a general discussion of the subject from different perspectives). For the results presented in the next section, a parameter set was accepted only if the steady-state ratio $(X_1 + X_2)/X_3 \leq 0.1$. This value for the ratio was chosen arbitrarily because there are no reliable measurements on which to base a more accurate estimate.

B2. Changes in the concentration of intermediates caused by changes in demand for the end product should be small. The previous condition ensures that the concentration of intermediates will not saturate the solvent capacity of a cell in a given steady state. However, if the metabolic conditions change and the demand for the end product of the pathway changes, this will cause the concentration of each intermediate to change, which may lead to saturation of the solvent capacity in the new steady state (e.g. Savageau, 1972). This could be prevented in our model if the absolute values of the logarithmic gains for intermediates, $|L(X_i, X_4)|$, are smaller than a predetermined value arbitrarily set at 0.5.

B3. Changes in the concentration of intermediates caused by changes in the initial substrate should be small. This will buffer the intermediate concentrations against changes in metabolism that are reflected in alterations in the level of initial substrate. This could be ensured if the absolute values of the logarithmic gains for intermediates, $|L(X_i, X_0)|$, are smaller than a predetermined value, e.g. $|L(X_i, X_0)| < 0.5\,(i = 1, 2, 3)$. This value is chosen arbitrarily because there are no reliable measurements on which to base a more accurate estimate.

B4. Systems should be robust, i.e. insensitive to spurious fluctuations in the parameters that define their structure (Savageau, 1972). We require that each intermediate have an aggregate sensitivity, defined as $\mathrm{SQRT}\left[\sum_j S(X_i, p_j)^2\right]$, less than a predetermine value arbitrarily set equal to 5.

B5. Each system should have a locally stable steady state. Systems without such stable steady states are dysfunctional because they are unable to maintain their homeostatic behavior in the face of spurious perturbations. The two margins of stability can be specified in terms of the last two Routh criteria (e.g. Savageau, 1976).

B6. Systems should have a rapid response time. This is related to the inverse of the turnover number (Dixon, 1958; Savageau, 1975), which should therefore be high. We require the turnover number for the pathway, defined as the pathway flux divided by the sum of the intermediate pools $(V/\sum_i X_i)$, to be larger than a predetermined value arbitrarily set equal to 1.

Results

Bias in the frequency distribution of parameter values

The values for each parameter were originally sampled with a uniform distribution. However, those parameter sets that define systems excluded from the composite behavioral class are rejected, and the frequency distribution of the accepted parameter values is therefore biased. The nature of the bias for each of the parameters can be determined from the histograms presented in Figure 2. We observe that the composite behavioral class has α_1 biased towards small values whereas α_2, α_3, and α_4 are biased towards large values. The kinetic order for the substrate of the pathway, g_{10}, is biased towards small values. Its frequency increases from $g_{10} = 0$ to $g_{10} = 0.3$ and then decreases exponentially until $g_{10} = 5$. A similar pattern is observed for g_{32}, although the frequency increases from 0 to 1.8, and then decrease but not exponentially. The kinetic order g_{21} is biased towards large values, and g_{43} is nearly uniform over its range. The inhibitory kinetic order for overall feedback, g_{13}, has a distribution with a central tendency, whereas the other inhibitory kinetic orders are almost uniformly distributed throughout their range of possible values.

We also determined the parameter distributions for each of the elementary behavioral classes (B1–B6 defined above) to see which, if any, might qualitatively reproduce the deviations from a uniform distribution that were observed for the composite behavioral class (Figure 2). Table 1 shows which elementary class is mainly responsible for the shape of each distribution in the composite behavioral class. In some cases, the distribution for the composite behavioral class can be attributed to the dominant influence of a particular elementary class (e.g. B3 in the case of g_{10}). In other cases, the distributions for the composite behavioral class can be attributed to the influence of several elementary classes acting in combination, which implies a synergistic influence (e.g. B1–B6 in the case α_3).

Frequency distribution for systemic properties of the ensemble

The frequency distributions for all steady-state properties of our model have long tails. These tails make it difficult to present informative histograms for each of the systemic

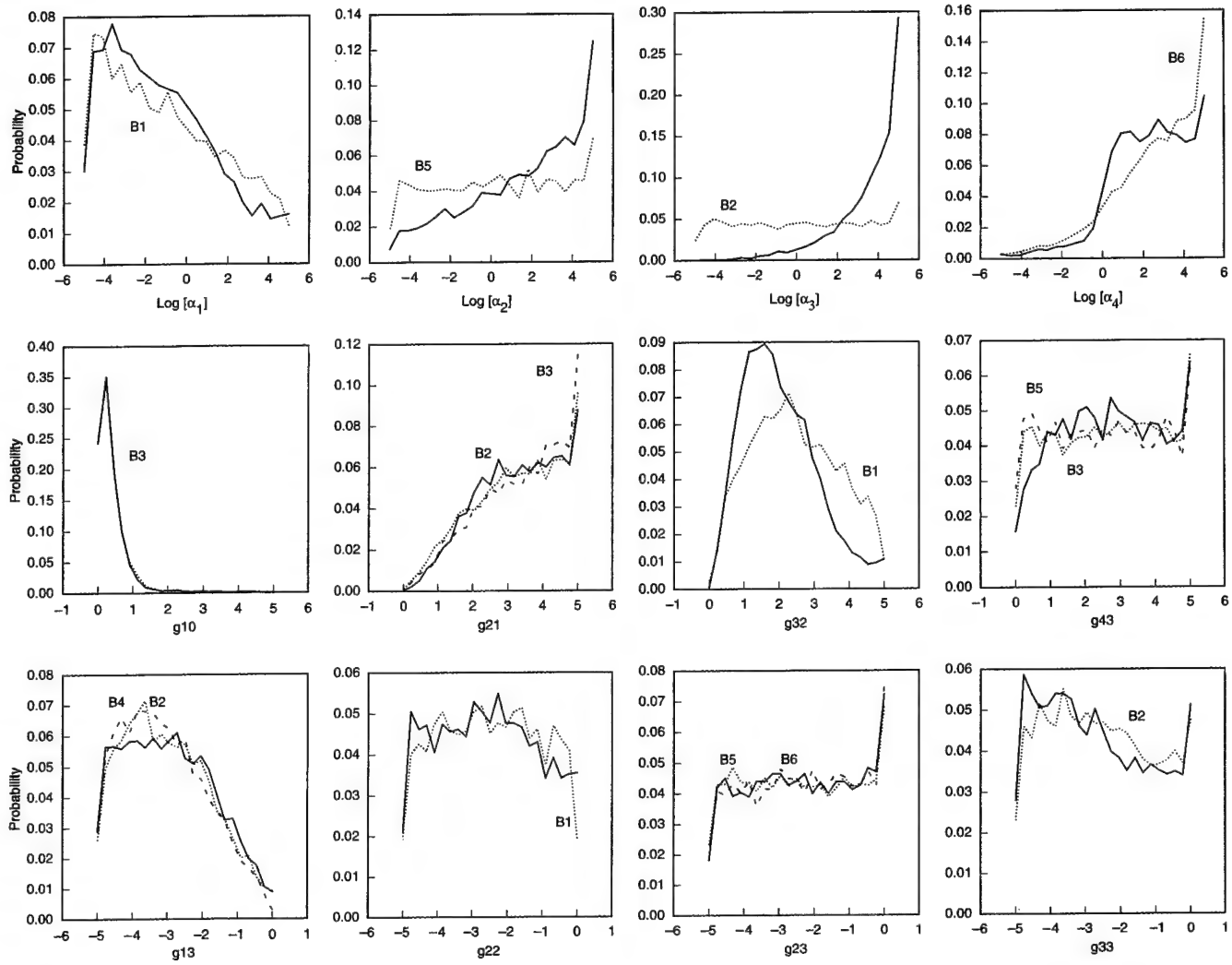


Fig. 2. Distribution of parameter values in ensembles of systems selected on the basis of various behavioral classes. Selection involved each of the six elementary behavioral classes (B1–B6) considered separately and the composite class consisting of all six elementary classes considered together. The solid line in each panel is the distribution for the composite behavioral class. Three different patterns are represented. In most cases the distribution for the composite class is closely represented by the distribution for one of the elementary classes (α_1, α_2, α_4, g_{10}, g_{22}, g_{33}). The distributions for the other elementary classes have very different shapes and are not shown. In four cases the distribution for the composite class is closely resembled by two or more of the distributions for the elementary classes (g_{21}, g_{43}, g_{13}, g_{23}). Distributions for only two of the elementary classes are shown. In two cases none of the distributions for the elementary classes is a close match to the distribution for the composite class (α_3, g_{32}). In these cases we show only the distribution for the elementary class that most closely resembles the distribution for the composite class.

properties. We chose to cut off the tails and add their frequency to the more extreme classes presented in the histograms. The results in this section are shown as histograms in Figure 3. We did not include histograms for the elementary behavioral classes because, in most cases, they have extremely long tails.

Steady-state concentrations and flux. All steady-state concentrations have frequency distributions that decrease as the concentration increases. At low concentrations the frequency decreases very sharply as the concentrations increase, but then the decrease becomes very small and there is a long tail in the distribution. The modal class for all of the frequency distributions is small. For X_1 and X_2 the modal class is in the interval [0,0.2334] with 90% and 75% of all systems in this interval, respectively. The modal class for X_3 has a larger value, in the interval [0.234,0.468]. Also, only 10% of all systems fall within

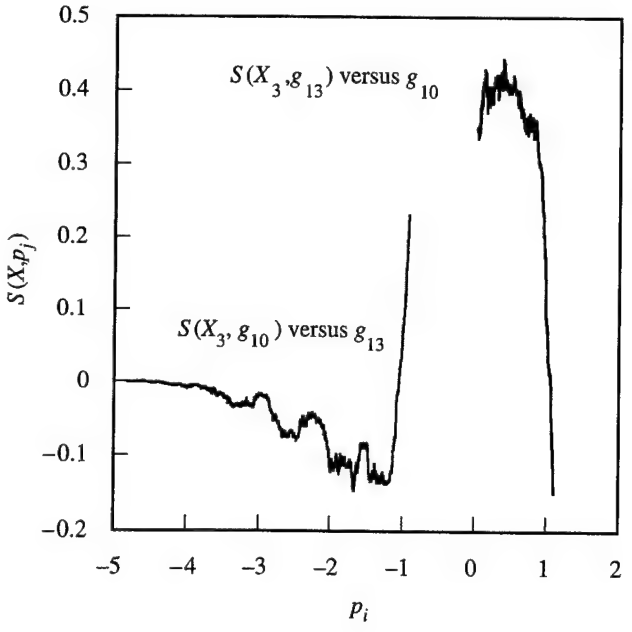


Fig. 5. Examples of graphs showing the statistical synergism between two different parameters. Each statistical synergism is determined by a plot of the moving median for the sensitivity with respect to parameter p_i versus the moving median for parameter p_j constructed from an ensemble of systems selected on the basis of the composite behavioral class. Note the asymmetry in the synergisms. See text for discussion.

Discussion

The study of generic biochemical systems requires a mathematical formalism that is systematically structured and capable of representing rather arbitrary nonlinear phenomena. The power-law formalism provides such a canonical nonlinear representation (Savageau, 1996), and for the work presented in this paper we have focused on the local S-system representation within this formalism. This representation, although nonlinear, has closed-form solutions for the steady state, and these can be used to study systemic properties analytically. However, more often than not, the complexity of the solutions for the properties of interest makes it difficult to analyze systemic behavior without assigning specific values to the parameters. In most cases these values are unknown; when they are known, they limit the interpretation of the results to a specific system and thus prevent generalization of the results. To overcome these limitations statistical studies involving large ensembles of random systems have been performed in a variety of contexts (see, e.g. Bhattacharjya and Liang, 1996; Glass, 1975; Kauffman, 1969a,b, 1993, and references therein). However, to our knowledge this approach has not been applied to continuous models for specific classes of biochemical systems with the objective of providing an exhaustive statistical characterization of their systemic properties.

In this work we have created large ensembles of randomly generated parameter values for a given structural class of biochemical systems and imposed selection on the basis of particular systemic properties. We then examined the resulting systems for bias in parameter values, bias in unselected systemic properties, and correlations among all their systemic properties.

Selection can be expected to influence the range of parameter values in the resulting systems. Although specific systemic properties have been used for some time as criteria to evolve networks towards optimality (e.g. James *et al.*, 1999), few, if any, attempts have been made to characterize the bias in parameter values that results from such a selection procedure. In fact, the usual view on the subject is that parameter values determine systemic behavior. We have had to take the opposite view to learn how selection for particular systemic behaviors influences the frequency distribution of parameter values. As seen in Figure 2, there are regular patterns of deviation from what was a uniform distribution before imposing selection based on the composite behavioral class. By using each of the elementary behavioral classes as an independent selection criterion we were able to determine whether any given elementary class made a major contribution to the observed bias in the distribution of any given parameter value. In some cases this is true (B3 in the case of g_{10}), whereas in others the distribution of parameter values for the composite class is the result of interplay among different elementary classes (B1–B6 in the case α_3).

Information on the distribution of parameter values is of interest in the design of experiments to measure the parameters in actual systems. By knowing the most probable values of a parameter, one can design experiments to target that range. Also, the use of behavioral classes to study specific kinds of systems provides an effective way to identify the relative importance of various regions in the parameter space of fit systems.

Selection for a particular systemic property may also influence other unselected systemic properties. As seen in Figure 3, selection on the basis of the composite behavioral class produces a frequency distribution for the values of the different systemic properties that is skewed in nearly every case, with a peak at low values and a long tail that decreases in frequency almost exponentially. (This is true of the distributions for the aggregate sensitivities, although it is not evident from the curves for X_1 and X_2 because their tails are off the scale.) The exceptions to this general pattern are the distributions for the logarithmic gain $L(X_1, X_0)$, which is nearly uniform over the range [0,5], and the logarithmic gains $L(X_1, X_4)$, $L(X_2, X_4)$, and $L(V, X_4)$, which exhibit a symmetric central tendency. We have also determined the

influence of the elementary behavioral classes on these distributions, but the results are less straightforward to interpret. In many cases the distribution for the composite class is quite different from the distribution for any of the elementary classes (data not shown). This indicates a strong synergism between different constraints that determine the distribution of values for the systemic properties.

Selection also can be expected to influence the correlations among the various systemic properties in the resulting systems. We have used a moving median technique (Alves and Savageau, 2000) to determine average correlations between different systemic properties. We found that these correlations exist and are, at least in some cases, dependent on the behavior of the system (Table 2). For example, most aggregate sensitivities are correlated with the concentrations by a symmetric curve of type C8 if no restrictions are imposed on the values of the concentrations in the system (data not shown). However, when we imposed the condition that intermediate concentrations be small (B1), this kind of symmetry breaks down (curves of type C8 and C9 become C3 and C4), because the systems being studied include only those with concentrations that have low values. It is important to note that, as a concentration tends to unity, the sensitivities to the kinetic orders associated with that concentration will tend to zero (due to the properties of the power law in the S-system formalism). This will tend to diminish the aggregate sensitivities that include these kinetic-order sensitivities. Table 2 also shows that in the composite class, robustness of intermediates and stability margins are inversely correlated; systems that have large stability margins have, on average, intermediates with high aggregate sensitivities and are thus less robust.

Finally, the same technique used to determine the correlations between different systemic properties also was used to determine the statistical synergisms between different parameters. The system in Figure 1 has small synergisms for the end product and flux (Table 3), because in many cases (54 out of 120) the sensitivities are not correlated with any parameter (statistical synergism is zero). Thus, the end product and the steady-state flux of the system are, on average, well buffered against second-order perturbations.

The approach illustrated in this paper provides statistical insights. It might be argued that biological systems are optimized and atypical, and thus not compatible with the application of statistical techniques. However, this objection is avoided in our approach. By defining behavioral classes for optimized systems, we are able to study the average behavior of optimized systems and not just the average behavior of random systems.

The methods we have described can in principle be applied to systems with more complex behaviors. For example, suppose we wish to consider biochemical systems that are capable of exhibiting either a single locally stable steady state (nongrowing cells that are viable but quiescent) or a single stable limit cycle (growing cells with a well-defined cycle time), depending only upon the value of an environmental cue. The behavioral classes that we would define for such systems would now include the combined properties of these two different modalities as well as the properties that might be applied to each of the separate modalities. The more complex behavioral class would include a number of dynamic properties (e.g. the period, amplitude, phase, and robustness of the oscillation, and the bifurcation value of the environmental cue for switching between modalities), and the analysis necessary to identify and characterize these behavioral classes would accordingly become more complex. The local S-system representation is capable of describing each of the separate modalities (Lewis, 1991), but not the two of them together with a given set of parameter values. For this purpose we would need the generalized-mass-action representation within the power-law formalism. This representation does not have analytical solutions for the steady state, and so the analysis and comparison of these properties would have to be done by numerical methods. Randomly generated sets of parameter values (which would now include values for a parameter representing the environmental cue) could be generated as before. However, we would now select only those sets of parameter values that satisfy the more complex behavioral class that includes both modalities and the appropriate switching between them in response to the environmental cue. Those sets of parameter values that only yield one of the two modalities would be excluded from consideration. This would ensure that any averaging procedure that is subsequently applied to systems with the randomly generated parameter sets would range over a homogeneous class of systems.

The approach proposed in this work also may be useful in providing information about systems that are poorly characterized. For example, suppose we know the structure of a system, but we are able to determine experimentally only some of the characteristic behaviors of the system. To be more specific, suppose we know that the concentrations of the system are within a given range, that increasing the value of a given independent variable will always cause a decrease in the values for some dependent variables, and that we are able to measure the range of values for the turnover times of the concentrations. With this information, we could generate ensembles of systems with the described characteristics and study them statistically. The results would allow us to make predictions about other systemic properties that might be measured and, for the unknown parameters,

about the range of values most likely to generate systems with the known behavior.

A combination of approaches will surely be needed to advance our understanding of large and complex systems in biology. We need to take advantage of the broad-scale capabilities of the top-down genomic technologies and the structural constraints provided by the more traditional bottom-up methodologies of molecular biology. We also need to identify the systemic regularities that exist even in randomly constructed networks. The approach presented in this paper appears well suited for the determination of such regularities in continuous models. It may facilitate the design of experiments to measure parameters by the bottom-up approach as well as provide a suitable framework to determine classes of models that give a good fit to data obtained by the top-down approach.

Acknowledgments

This work was supported in part by a joint PhD fellowship PRAXIS XXI / BD / 9803 / 96 granted by PRAXIS XXI through Programa Gulbenkian de Doutoramentos em Biologia e Medicina (RA), the US Public Health Service Grant RO1-GM30054 from the National Institutes of Health (MAS), and the US Department of Defense Grant N00014-97-1-0364 from the Office of Naval Research (MAS.). We thank Armindo Salvador for critically reading early versions of this manuscript and making useful comments.

References

Albe,K.R. and Wright,B.E. (1992) Systems analysis of the tricarboxylic acid cycle in Dictyostelium discoideum. II. Control analysis. *J. Biol. Chem.*, **267**, 3106–3114.

Alves,R. and Savageau,M.A. (2000) Comparing systemic properties of ensembles of biological networks by graphical and statistical methods. *Bioinformatics*, **16**, 527–533.

Antunes,F., Salvador,A., Marinho,H.S., Alves,R. and Pinto,R.E. (1996) Lipid peroxidation in mitochondrial inner membranes 1. An integrative kinetic model. *Free Rad. Biol. Med.*, **21**, 917–943.

Atkinson,D.E. (1969) Limitation of metabolite concentrations and the conservation of solvent capacity in the living cell. *Curr. Top. Cell. Regulation*, **1**, 29–43.

Bhattacharjya,A. and Liang,S. (1996) Median attractor and transients in random Boolean nets. *Physica*, D **95**, 29–34.

Bliss,R.D., Painter,P.R. and Marr,A.G. (1982) Role of feedback inhibition in stabilizing the classical operon. *J. Theoret. Biol.*, **97**, 177–193.

Bratley,P. and Fox,B.L. (1988) Implementing Sobol quasirandom sequence generator. *ACM Trans. Math. Software*, **14**, 88–100.

Brown,P.O. and Botstein,D. (1999) Exploring the new world of the genome with DNA microarrays. *Nat. Genet.*, **21**, 33–37.

Brown,G.C., Hafner,R.P. and Brand,M.D. (1990) A top-down approach to the determination of control coefficients in metabolic control theory. *Eur. J. Biochem.*, **188**, 321–325.

Chappell,M.J. and Godfrey,K.R. (1992) Structure identifiability of the parameters of a non linear batch reactor model. *Math. Biosci.*, **108**, 241–251.

Chu,S., DeRisi,J.L., Eisen,M., Mulholland,J., Botstein,D., Brown,P.O. and Herskowitz,I. (1998) The transcriptional program of sporulation in budding yeast. *Science*, **282**, 699–705.

Clegg,J.S. (1984) Properties and metabolism of the aqueous cytoplasm and its boundaries. *Am. J. Physiol.*, **246**, R133–R151.

Curto,R., Voit,E.O., Sorribas,A. and Cascante,M. (1998) Mathematical models of purine metabolism in man. *Math Biosci.*, **151**, 1–49.

DeRisi,J.L., Iyer,V.R. and Brown,P.O. (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science*, **278**, 680–686.

Diamond,J.M. (1975) How do biological systems discriminate among physically similar ions? *J. Exp. Zool.*, **194**, 227–239.

Dixon,M. (1958) *Enzymes*. Academic Press, New York.

Domnitz,R.H. (1976) Lateral position and interaural discrimination—data and black box model. *J. Acoust. Soc. Am.*, **59**, S23.

Dunn,O.J. and Clark,V.A. (1974) Applied statistics: analysis of variance and regression. Wiley, New York.

Eisen,M.B., Spellman,P.T., Brown,P.O. and Botstein,D. (1998) Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA*, **95**, 14863–14868.

Feng,D.G. and DiStefano,J.J. (1991) Cut set analysis of compartmental-models with applications to experimental design. *Am. J. Physiol.*, E284–E269.

Ginn,T.R. and Cushman,J.H. (1992) A continuous time inverse operator for groundwater and contaminant transport modeling—model identifiability. *Water Resour. Res.*, **28**, 539–549.

Glass,L. (1975) Classification of biological networks by their qualitative dynamics. *J. Theoret. Biol.*, **54**, 85–107.

Hlavacek,W.S. and Savageau,M.A. (1995) Subunit structure of regulator proteins influences the design of gene circuitry: analysis of perfectly coupled and completely uncoupled circuits. *J. Mol. Biol.*, **248**, 739–755.

James,A., Swann,K. and Recce,M. (1999) Cell behaviour as a dynamic attractor in the intracellular signalling system. *J. Theoret. Biol.*, **196**, 269–288.

Kargi,F. and Shuler,M.L. (1979) Generalized differential specific rate equation for microbial growth. *Biotech. Bioeng.*, **21**, 1871–1875.

Kauffman,S.A. (1969a) Metabolic stability and epigenesis in randomly connected nets. *J. Theoret. Biol.*, **22**, 437–467.

Kauffman,S.A. (1969b) Homeostasis and differentiation in random genetic control networks. *Nature*, **224**, 177–178.

Kauffman,S.A. (1993) *The Origins of Order: Self-Organization and Selection in Evolution*. Oxford University Press, New York.

Levine,R.L. and Ginsburg,A. (1985) Modulation by molecular interactions. *Curr. Top. Cell. Reg.*, **26**, 1–549.

Lewis,D.C. (1991) Qualitative analysis of S-systems: Hopf bifurcations. In Voit,E.O. (ed.), *Canonical Nonlinear Modeling: S-Systems Approach to Understanding Complexity* Van Nostrand Reinhold, New York, pp. 304–344.

Moheimani,S.O. R., Savkin,A.V. and Petersen,I.R. (1996) Robust observability for a class of time-varying discrete-time uncertain systems. *Syst. Control Lett.*, **27**, 261–266.

Moore,G.E., Gadol,S.M., Robinson,J.B. and Srere,P.A. (1984)

Binding of citrate synthase and malate-dehydrogenase to mitochondrial inner membranes—tissue distribution and metabolic effects. *Biochem Biophys. Res. Comm.*, **121**, 612–618.

Ni,T.-C. and Savageau,M.A. (1996) Model assessment and refinement using strategies from biochemical systems theory: application to metabolism in human red blood cells. *J. Theoret. Biol.*, **179**, 329–368.

Ovadi,J. and Srere,P.A. (1996) Metabolic consequences of enzyme interactions. *Cell Biochem. Func.*, **14**, 249–258.

Quant,P. (1993) Experimental application of the top-down control analysis to metabolic systems. *Trends Biochem Sci.*, **18**, 26–30.

Salvador,A. (1999a) Synergism analysis of biochemical systems. I. Conceptual framework. *Math. Biosci.*, **163**, 105–129.

Salvador,A. (1999b) Synergism analysis of biochemical systems. II. Deviations from multiplicativity. *Math. Biosci.*, **163**, 131–158.

Savageau,M.A. (1969) Biochemical systems analysis II. The steady state solutions for an n-pool system using a power-law approximation. *J. Theoret. Biol.*, **25**, 370–379.

Savageau,M.A. (1971) Concepts relating the behavior of biochemical systems to their underlying molecular properties. *Arch. Biochem. Biophys.*, **145**, 612–621.

Savageau,M.A. (1972) The behavior of intact biochemical control systems. *Curr. Top. Cell Reg.*, **6**, 63–110.

Savageau,M.A. (1975) Optimal design of feedback control by inhibition: dynamical considerations. *J. Mol. Evol.*, **5**, 199–222.

Savageau,M.A. (1976) *Biochemical Systems Analysis: A Study of Function and Design in Molecular Biology*. Addison-Wesley, Reading, MA.

Savageau,M.A. (1992) A critique of the enzymologist's test tube. In Bittar,E.E. (ed.), *Fundamentals of Medical Cell Biology* Vol 3A, JAI Press Inc., Greenwich, Connecticut, pp. 45–108.

Savageau,M.A. (1996) A kinetic formalism for integrative molecular biology: manifestation in biochemical systems theory and use in elucidating design principles for gene circuits. In Collado-Vides,J., Magasanik,B. and Smith,T.F. (eds), *Integrative Approaches to Molecular Biology* MIT Press, Cambridge, MA, pp. 115–146.

Shiraishi,F. and Savageau,M.A. (1993) The tricarboxylic-acid cycle in *Dictyostelium discoideum*: systemic effects of including protein turnover in the current model. *J. Biol. Chem.*, **268**, 16917–16028.

Somogyi,R., Fuhrman,S., Askenazi,M. and Wuensche,A. (1997) The gene expression matrix: towards the extraction of genetic network architectures. *Nonlinear Anal. Theor.*, 1824–1815.

Sorribas,A. and Savageau,M.A. (1989) Strategies for representing metabolic pathways within biochemical systems theory: reversible pathways. *Math. Biosci.*, **94**, 239–269.

Sorribas,A., Samitier,S., Canela,E.I. and Cascante,M. (1993) Metabolic pathway characterization from transient-response data obtained *in situ*. Parameter estimation in S-system models. *J. Theoret. Biol.*, **162**, 81–102.

Srere,P (1987) Complexes of sequential metabolic enzymes. *Ann. Rev. Biochem.*, **56**, 89–124.

Törönen,P., Kolehmainen,M., Wong,G. and Castrén,E. (1999) Analysis of gene expression data using self-organizing maps. *FEBS Lett.*, **451**, 142–146.

Voit,E.O. and Savageau,M.A. (1982) Power-law approach to modeling biological systems II. Application to ethanol production. *J. Ferment. Technol.*, **60**, 229–232.

Voit,E.O. and Savageau,M.A. (1987) Accuracy of alternative representations for integrated biochemical systems. *Biochemistry*, **26**, 6869–6880.

Xu,L., Saito,O. and Abe,K. (1996) Design of practically stable *n*-dimensional feedback systems: a state-space approach. *Int. J. Control*, **64**, 29–39.

Yanofsky,C. and Horn,V. (1994) Role of regulatory features of the *trp* operon of *Escherichia coli* in mediating a response to a nutritional shift. *J. Bacteriol.*, **176**, 6245–6254.

Vol. 16 no. 9 2000
Pages 786–798

Extending the method of mathematically controlled comparison to include numerical comparisons

Rui Alves [1, 2, 3] and Michael A. Savageau [1,*]

[1]Department of Microbiology and Immunology, University of Michigan Medical School, 5641 Medical Science Building II, Ann Arbor, MI 48109-0620, USA, [2]Grupo de Bioquimica e Biologia Teoricas, Instituto Rocha Cabral, Calçada Bento da Rocha Cabral 14, 1250, Lisboa, Portugal and [3]Programa Gulbenkian de Doutoramentos em Biologia e Medicina, Departamento de Ensino, Instituto Gulbenkian de Ciência, Rua da Quinta Grande 6, 1800, Oeiras, Portugal

Received on December 12, 1999; revised on March 28, 2000; accepted on May 15, 2000

Abstract

Motivation: *The method of mathematically controlled comparison has been used for some time to determine which of two alternative regulatory designs is better according to specific quantitative criteria for functional effectiveness. In some cases, the results obtained using this technique are general and independent of parameter values and the answers are clear-cut. In others, the result might be general, but the demonstration is difficult and numerical results with specific parameter values can help to clarify the situation. In either case, numerical results with specific parameter values can also provide an answer to the question of how much larger the values might be. In contrast, a more ambiguous result is obtained when either of the alternatives can have the larger value for a given systemic property, depending on the specific values of the parameters. In any case, introduction of specific values for the parameters reduces the generality of the results. Therefore, we have been motivated to develop and apply statistical methods that would permit the use of numerical values for the parameters and yet retain some of the generality that makes mathematically controlled comparison so attractive.*

Results: *We illustrate this new numerical method in a step-by-step application using a very simple didactic example. We also validate the results by comparison with the corresponding results obtained using the previously developed analytical method. The analytical approach is briefly present for reference purposes, since some of the same key concepts are needed to understand the numerical method and the results are needed for comparison. The numerical method confirms the qualitative differences between the systemic behavior of alternative designs obtained from the analytical method. In addition, the numerical method allows for quantification of the differences and it provides results that are general in a statistical sense. For example, the older analytical method showed that overall feedback inhibition in an unbranched pathway makes the system more robust whereas it decreases the stability margin of the steady state. The numerical method shows that the magnitudes of these differences are not comparable. The differences in stability margins (1–2% on average) are small when compared to the differences in robustness (50–100% on average). Furthermore, the numerical method shows that the system with overall feedback responds more quickly to change than the otherwise equivalent system without overall feedback. These results suggest reasons why overall feedback inhibition is such a prevalent regulatory pattern in unbranched biosynthetic pathways.*

Contact: *savageau@umich.edu*

Introduction

The experimental investigation of biological regulatory mechanisms has revealed an enormous variety of alternative molecular designs and raised questions about their function, design and evolution. Mathematically controlled comparison is a technique that was specifically developed to study such alternative regulatory designs (Savageau, 1972). By using the mathematical analog of a well-controlled experiment, this technique analytically determines the irreducible qualitative differences in the systemic behavior of the alternative designs. This technique has been used to study alternative regulatory designs in metabolic pathways (e.g. Savageau, 1974; Hunding, 1974; Savageau and Jacknow, 1979), in gene circuits (e.g. Hlavacek and Savageau, 1996), in immune

*To whom correspondence should be addressed.

© Oxford University Press 2000

networks (e.g. Irvine and Savageau, 1985; De Boer and Hogeweg, 1989a,b), and in the host-pathogen response to HIV infection (De Boer and Perelson, 1998). The introduction of numerical values for the parameters provides quantification of these differences in specific cases but eliminates the generality of the results. In this paper we introduce a numerical approach to mathematically controlled comparisons that allows the introduction of specific numerical values for the parameters in the analysis while still retaining the generality of the results in a statistical sense.

The most common use of mathematically controlled comparison requires the existence of closed-form solutions for the steady state. Such solutions can be obtained by using the local S-system representation to characterize the pathway of interest. Important functional constraints are introduced by equating relevant steady-state properties of the alternative systems being compared. Further analysis (dynamic as well as steady state) is performed and a profile of ratios for corresponding results from the alternative systems is constructed. In some cases, a ratio can be determined analytically to be less than, equal to, or greater than unity. For example, if the ratio of values for property P in a reference system to the same property in an alternate system is larger than unity, then the reference system can always be made to have a larger value for P no matter how large the value for P in the alternate system.

However, if one wishes to know how much greater than unity a given ratio is, then one needs to examine actual values for the parameters. These parameter values are not always available or, if available, are not always accurate. Moreover, there are cases in which the ratio can be less than or greater than unity depending upon the specific values for the parameters. In any case, the results of such a numerical comparison are no longer general. In this work we propose a novel approach to this problem that combines the method of mathematically controlled comparison with statistical techniques (Alves and Savageau, 2000a,b) to yield numerical results that are general in a statistical sense.

Although we could describe the numerical method in general terms, this approach would be too abstract and difficult to understand. Instead, we will illustrate this new numerical method by means of a step-by-step application using a very simple didactic example. We also validate the results by comparison with the corresponding results obtained using the previously developed analytical method. The analytical approach is briefly presented for reference purposes, because some of the same key concepts are needed to understand the numerical method and because the results are needed for comparison.

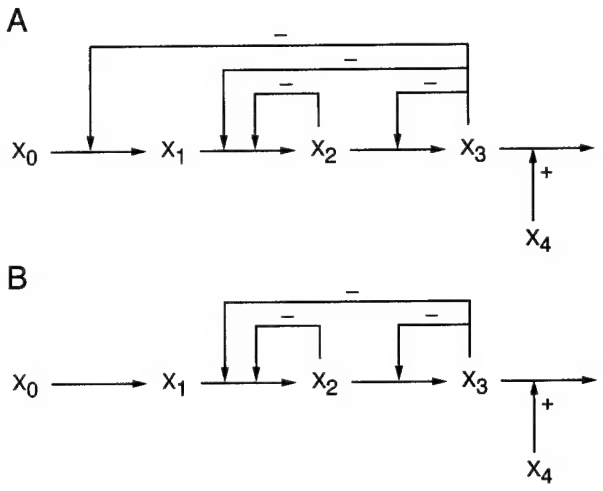


Fig. 1. Three-step unbranched biosynthetic pathways with inhibitory feedback. The metabolites in the pathways are represented by X_1 to X_3, and their concentrations are dependent variables. The initial substrate is represented by X_0 and the modifier of the demand for end product is represented by X_4; the concentrations of these latter metabolites are considered independent variables. The horizontal arrows represent chemical conversion, whereas the vertical arrows represent regulatory influences. This pathway can be viewed as an abstraction of biosynthetic pathways for amino acids. (A) Pathway with overall feedback inhibition (reference model). (B) Pathway without overall feedback inhibition (alternative model).

Methods

Alternative models

The didactic example that we use to illustrate our numerical method is an unbranched three-step pathway as shown schematically in Figure 1. This is an abstraction from actual three-step biosynthetic pathways such as those involved in the biosynthesis of amino acids (e.g. http://www.genome.ad.jp/kegg/dblinks/map/map01150.html). The independent variable X_4 represents the cell's demand for the end product X_3. If the cell requires large amounts of X_3, then the value of X_4 will be high; if small amounts of X_3 are required, then the value of X_4 will be low. These models show the pathway with and without end-product inhibition (Umbarger, 1956; Yates and Pardee, 1956; Monod *et al.*, 1963), a common feature of such pathways. We have observed (by consulting the database at http://wit.mcs.anl.gov//EMP/) that there is usually no other feedback to the first step of the pathway. However, feedback to intermediate reactions may exist and for this reason we consider models with all possible intermediary feedback interactions.

Differential equations

The dynamic behavior of each model can be described by a set of ordinary differential equations, one equation per

intermediate. This set of equations can be approximated to the first order in logarithmic space, yielding the canonical form for the local S-system representation (Savageau, 1969, 1996). For the model in Figure 1a this equation set becomes:

$$\begin{aligned} dX_1/dt &= \alpha_1 X_0^{g10} X_3^{g13} - \alpha_2 \prod_{j=1}^{3} X_j^{g2j} \\ dX_2/dt &= \alpha_2 \prod_{j=1}^{3} X_j^{g2j} - \alpha_3 \prod_{j=2}^{3} X_j^{g3j} \\ dX_3/dt &= \alpha_3 \prod_{j=2}^{3} X_j^{g3j} - \alpha_4 X_3^{g43} X_4^{g44} \end{aligned} \tag{1}$$

For the model in Figure 1b the equation set becomes:

$$\begin{aligned} dX_1/dt &= \alpha_1' X_0^{g'10} - \alpha_2 \prod_{j=1}^{3} X_j^{g2j} \\ dX_2/dt &= \alpha_2 \prod_{j=1}^{3} X_j^{g2j} - \alpha_3 \prod_{j=2}^{3} X_j^{g3j} \\ dX_3/dt &= \alpha_3 \prod_{j=2}^{3} X_j^{g3j} - \alpha_4 X_3^{g43} X_4^{g44} \end{aligned} \tag{2}$$

The multiplicative parameters (rate constants), α, influence the time scales of the reactions and are always positive. The exponential parameters (kinetic orders), g, represent the influence of each metabolite on each aggregate rate law. If X_i influences the aggregate rate law V_j, either as a substrate or a modulator, and if an increase in the concentration of X_i causes an increase in the rate V_j, then the kinetic order will be positive. If an increase in the concentration of X_i causes a decrease in the rate V_j, then the kinetic order will be negative. If an increase in the concentration of X_i causes neither an increase nor a decrease in the rate V_j, then the kinetic order will be zero. Thus, the positive kinetic orders in Equation (1) are $g_{i+1,i}$ $(0 \leq i \leq 3)$, which are the kinetic orders for substrates of reactions, and g_{44}, which is a scale factor arbitrarily set equal to 1.0. The remaining kinetic orders are negative, since these represent negative feedback interactions.

The temporal responsiveness of each model can be determined by perturbing the system variables, solving the corresponding dynamic equations, and calculating the time for the dependent variables to settle within 1% of their final steady-state values.

Steady-state solution and key systemic properties

At the steady state, which can be analytically determined, both the production and consumption terms have identical values. One can write the following matrix equation (Savageau, 1969):

$$\begin{bmatrix} b_1 - g_{10}Y_0 \\ b_2 \\ b_3 + Y_4 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix}$$

$$\mathbf{B} = \mathbf{AY} \tag{3}$$

where $b_i = \ln(\alpha_{i+1}/\alpha_i)$, $a_{ij} = g_{ij} - g_{i+1,j}$ and $Y_i = \ln(X_i)$. Equation (3) is linear and therefore easily solved to obtain the steady-state values for each Y_i; the corresponding values for each X_i are then obtained by simple exponentiation.

Two types of coefficients, logarithmic gains and parameter sensitivities, can be used to characterize the steady state of such models (Savageau, 1971). Logarithmic gains measure the relative influence of each independent variable on each dependent variable of the model. For example,

$$L(X_i, X_0) = \frac{d\,\mathrm{Log}(X_i)}{d\,\mathrm{Log}(X_0)} = \frac{dY_i}{dY_0} \tag{4}$$

measures the percent change in the concentration of intermediate X_i caused by a percentage change in the concentration of the initial substrate X_0. Logarithmic gains provide important information concerning the amplification or attenuation of signals as they are propagated through the system.

Parameter sensitivities measure the relative influence of each parameter on each dependent variable of the model. For example,

$$S(X_i, p_j) = \frac{d\,\mathrm{Log}(X_i)}{d\,\mathrm{Log}(p_j)} = p_j \frac{dY_i}{dp_j} \tag{5}$$

measures the percent change in the concentration of intermediate X_i caused by a percentage change in the value of the parameter p_j. Parameter sensitivities provide important information about system robustness, i.e. how sensitive the system is to perturbations in the parameters that define the structure of the system.

Since steady-state solutions exist in closed form we can calculate each of the two types of coefficients simply by taking the appropriate derivatives. Although the mathematical operations involved are the same in each case, it is important to keep in mind that the biological significance of the two types of coefficients is very different.

The local stability of the steady state can be determined by applying the Routh criteria (Dorf, 1992). The magnitude of the two critical Routh conditions can be used to quantify the margin of stability (Hlavacek and Savageau, 1996).

Responsiveness

Systems should respond quickly to changes in their environment. To evaluate temporal responsiveness, perturbations of 20% were made in the steady-state values of the intermediate concentrations and the time required for them to settle within 1% of their final steady-state values was then calculated. This also gives an indication of the transit time for metabolites in the system. These transit times should be low. There is no exact way to determine the transient time analytically. Thus, this part of the analysis will be dealt with only in the numerical section of the results.

Generation of random ensembles

The analytical results give qualitative information that characterizes the role of overall feedback in the system of Figure 1A. To obtain quantitative information, one must introduce specific values for the parameters and compare systems. For this purpose we have randomly generated a large ensemble of parameter sets and selected 5000 of these sets that define systems consistent with various physical and biochemical constraints. These constraints include mass balance, low concentrations of intermediates and small changes in their value to minimize the utilization of the solvent capacity in the cell, small values for parameter sensitivities so as to desensitize the system to spurious fluctuations affecting its structure, and stability margins large enough to ensure local stability of the systems. A detailed description of these methods can be found in Alves and Savageau (2000b). Mathematica™ (Wolfram, 1997) was used for all numerical procedures.

Density of ratios plot

To interpret the ratios that result from our analysis we use Density of Ratios plots as defined in Alves and Savageau (2000a). The primary density plots from the raw data have the magnitude for some property of the reference model on the x-axis and the corresponding ratio of magnitudes (reference model to alternative model) on the y-axis. The primary plot can be viewed as a list of 5000 paired values that can be ordered with respect to the reference magnitude to form a list L_1 in which the first pair has the lowest measured value for property P in the reference model, the second has the second lowest, and so on. Secondary density plots are constructed from the primary plots by the use of moving quantile techniques with a window size of 500. The procedure is as follows. One collects the first 500 ratios from the list L_1, calculates the quantile of interest for this sample, and pairs this number $\langle R \rangle$, with the median value of the corresponding P values of the reference model denoted $\langle P \rangle$. One advances the window by one position, collects ratios 2 to 501, calculates $\langle R \rangle$, and pairs it with the corresponding $\langle P \rangle$ value and continues in this manner until the last ratio from the list L_1 was used for the first time (for further explanation of moving median techniques see, e.g. Hamilton, 1994). The slope in the secondary plot measures the degree of correlation between the quantities plotted on the x- and y-axes. This technique is also used to examine correlations between ratios of interest and other magnitudes shared by the two systems, e.g. the correlation between the ratio of stability margins and the magnitude of a rate constant common to the two systems (for traditional applications of correlation analysis see Wherry, 1984).

Analytical comparison

Firstly, we shall exemplify the analytical aspects of a mathematically controlled comparison aimed at discovering the advantages, if any, brought about by overall feedback inhibition. This will serve to introduce key concepts that will be needed for the numerical aspects in the following section. Also, the results will be used for later comparison to validate the results obtained with the new numerical method.

We compare the systemic behavior of the model in Figure 1A (reference model) with that in Figure 1B (alternative model). To ensure that the results are due solely to the differences in design and not reversible by a mere change in parameter value, we shall insist on the following mathematical controls.

Internal and external equivalence

Only the first step in the pathway is allowed to differ between the reference model and the alternative. Therefore, to establish an internal equivalence (Savageau, 1972, 1976; Irvine, 1991) between the two designs, we require the values for the corresponding parameters of all other steps in the two models to be the same.

The first step of the pathway differs between the reference model and the alternative. If we reason that loss or gain of an inhibitory site on the first enzyme comes about by mutation, and that this mutation can cause changes in all the parameters of the process, then (taking the model in Figure 1A as reference) a mutation causing loss of overall feedback inhibition would change the parameters g_{10}, g_{13} and α_1 in Equation (1) to g'_{10}, $g'_{13} = 0$ and α'_1 in Equation (2). Since we wish to determine the effects that are due solely to changes in the structure of the system, we shall specify new values for g'_{10} and α'_1 that minimize all other effects. This can be accomplished by deriving the mathematical expression for a given steady-state property in each of the two models, equating these expressions, and then solving the constraint equation for the value of a primed parameter. For example, if we derive expressions for the logarithmic gains and require that $L(X_i, X_0)_A = L(X_i, X_0)_B$, then this equation can be solved to determine the following value for g'_{10}.

$$g'_{10} = \frac{g_{10}g_{43}}{g_{43} - g_{13}} \tag{6}$$

Similarly, we can derive expressions for the steady-state concentrations and require that $[X_i]_A = [X_i]_B$, then this equation can be solved to determine the following value for α'_1.

$$\text{Log}[\alpha'_1] = \frac{g_{43}\,\text{Log}[\alpha_1] + g_{13}\,\text{Log}[\beta_3]}{g_{43} - g_{13}} \tag{7}$$

These particular values for the primed parameters make the steady-state flux, each of the corresponding steady-state concentrations, and the logarithmic gains in each of these quantities the same in both models. The process we have just described determines the maximal degree of external equivalence between the two models. There are no more 'free' parameters that can be used to reduce the differences and all remaining differences can be attributed to the change in system structure, i.e. to overall feedback inhibition. The external equivalency conditions we require insure that both the reference and the alternative models have the same steady-state concentrations and rates. This implies that the steady-state thermodynamic potential across each corresponding reaction is the same in the reference and the alternative models. Having established the conditions for maximal equivalence, we can now analyze the two models and determine their remaining differences.

Pathway gain

The logarithmic gains in concentrations and fluxes with respect to changes in the initial substrate X_0 determine whether the pathway is amplifying or homeostatic. When comparing pathways designed to amplify biochemical signals it is important that the alternatives provide the same high logarithmic gain. Conversely, when comparing pathways designed to attenuate biochemical signals it is important that the alternatives have the same low logarithmic gain. The method of mathematically controlled comparison insures that both the compared models will have the same logarithmic gains and thus have the same amplifying or homeostatic characteristics.

The logarithmic gains in concentration with respect to changes in demand (represented here by changes in the modulator X_4) are smaller in the reference system, whereas the logarithmic gain in flux with respect to changes in X_4 is larger in the reference system. In this aspect, the reference system is more efficient because it can produce greater increases in flux with smaller increases in concentration. When the logarithmic gain in flux with respect to changes in demand is zero (as is the case in the alternative model), changes in demand have no influence over the flux. Thus, overall feedback inhibition makes the system better equipped to deal with changes in the demand for X_3. These results are shown in Table 1.

Table 1. Analytical expressions for the ratios of corresponding systemic properties of the reference system and the alternative system

Systemic property	Dependent variable of the system			
	X_1	X_2	X_3	V
$L(\cdot, X_0)$	1	1	1	1
$L(\cdot, X_4)$	B^a	C	A	$1/0^b$
$S(\cdot, \alpha_1)$	A	A	A	A
$S(\cdot, \alpha_2)$	1	1	1	1
$S(\cdot, \alpha_3)$	1	1	1	1
$S(\cdot, \alpha_4)$	B	C	A	1/0
$S(\cdot, g_{10})$	1	1	1	1
$S(\cdot, g_{13})$	--	--	--	--
$S(\cdot, g_{21})$	1	1	1	1
$S(\cdot, g_{22})$	1	1	1	1
$S(\cdot, g_{23})$	1	1	1	1
$S(\cdot, g_{32})$	1	1	1	1
$S(\cdot, g_{33})$	1	1	1	1
$S(\cdot, g_{43})$	B	C	A	1/0

[a] The three critical ratios are given by the following analytical expressions:

$$A = 1 + \frac{g_{13}}{g_{43} - g_{13}} < 1$$

$$B = 1 + \frac{g_{13}[g_{22}(g_{33} - g_{43}) + g_{32}(g_{43} - g_{23})]}{(g_{13} - g_{43})(g_{23}g_{32} - g_{22}g_{33})} < 1$$

$$C = 1 + \frac{g_{13}(g_{43} - g_{33})}{g_{33}(g_{13} - g_{43})} < 1$$

[b] The ratio 1/0 represents the division of any non-zero number by zero.

Robustness

The system should be robust, i.e. insensitive to fluctuations in the parameter values (Shiraishi and Savageau, 1992). This means that the sensitivity profile should, in general, be as low as possible. Whenever the sensitivity of a concentration to a parameter is different in the two models, it is smaller in the reference model, i.e. the ratio $S(X_i, p_j)_A / S(X_i, p_j)_B$ is always less than or equal to unity. Thus, overall feedback inhibition makes each intermediate concentration less sensitive (i.e. more robust) with respect to fluctuations in parameter values.

Most of the corresponding flux sensitivities are equal in the reference and alternative models. The sensitivity $S(V, \alpha_1)$ is smaller in the reference model, which makes this model less sensitive to changes in the molecular activity of the first enzyme, whereas the sensitivities $S(V, g_{43})$ and $S(V, \alpha_4)$ are larger in the reference system because they also can reflect changes in demand. These results are shown in Table 1.

Stability

The steady state of the system should be stable, i.e. the system should return to its original steady state after a small perturbation. If this does not occur, the system is dysfunctional. The margins of stability for a system can be measured using the Routh criteria (Hlavacek and

Savageau, 1996). The larger these margins, the further from the boundaries of instability the system will be. The results of the analysis are as follows:

$$\frac{\text{Criterion\#1}_A}{\text{Criterion\#1}_B} = \frac{\text{Criterion\#2}_A}{\text{Criterion\#2}_B} = 1 \quad \frac{\text{Criterion\#3}_A}{\text{Criterion\#3}_B} = D < 1 \tag{8}$$

where

$$D = 1 + \frac{F_1F_2F_3g_{13}g_{21}g_{32}}{\left(\begin{array}{l} F_1^2F_2g_{21}^2g_{32} - F_1F_2^2g_{21}g_{22}g_{32} \\ -F_1F_2F_3g_{21}g_{23}g_{32} + F_2^2F_3g_{22}g_{23}g_{32} \\ +F_1F_2^2g_{21}g_{32}^2 - F_2^2F_3g_{23}g_{32}^2 \\ -F_1F_3^2g_{21}^2g_{33} + 2F_1F_2F_3g_{21}g_{22}g_{33} \\ -F_2^2F_3g_{22}^2g_{33} - 2F_1F_2F_3g_{21}g_{32}g_{33} \\ +F_2^2F_3g_{22}g_{32}g_{33} + F_2F_3^2g_{23}g_{32}g_{33} \\ +F_1F_3^2g_{21}g_{33}^2 - F_2F_3^2g_{22}g_{33}^2 \\ +F_1^2F_3g_{21}^2g_{43} - 2F_1F_2F_3g_{21}g_{22}g_{43} \\ +F_2^2F_3g_{22}^2g_{43} + 2F_1F_2F_3g_{21}g_{32}g_{43} \\ -2F_2^2F_3g_{22}g_{32}g_{43} - F_2F_3^2g_{23}g_{32}g_{43} \\ +F_2^2F_3g_{32}^2g_{43} - 2F_1F_3^2g_{21}g_{33}g_{43} \\ +2F_2F_3^2g_{22}g_{33}g_{43} - F_2F_3^2g_{32}g_{33}g_{43} \\ +F_1F_3^2g_{21}g_{43}^2 - F_2F_3^2g_{22}g_{43}^2 \\ +F_2F_3^2g_{32}g_{43}^2 \end{array}\right)}$$

with F_i being the turn-over number of the pool X_i. Note that the negative signs in this expression always precede parameters that represent negative feedback and consequently all terms have positive values. Thus, the reference and the alternative models differ in only one of the three Routh criteria applicable for a three variable system, and the alternative system has the larger margin of stability.

Summary

The analytical comparison gives qualitative results that characterize the role of overall feedback inhibition in the model of Figure 1A. This analysis demonstrates that the model with overall feedback inhibition is more robust and that its flux is more responsive to changes in demand for the end product, although this model has a smaller margin of stability. However, this analysis does not tell us how much more robust or how much more responsive to demand the reference model is, nor does it tell us how much smaller its margin of stability is. For answers to these questions we must consider specific values for the parameters and employ statistical techniques if we are to uncover general tendencies.

Numerical comparison

The techniques described in Alves and Savageau (2000b) have been used to generate an ensemble of 5000 parameter sets that characterize and reference the alternative systems with stable steady states. Each of these parameter sets was then inserted into the appropriate equations to determine the magnitude of the quantitative differences between reference and alternative systems.

Ratios of systemic properties

Figure 2A shows a typical Density of Ratios plot for an individual parameter sensitivity. One can clearly see that $S(X_i, p_j)_A/S(X_i, p_j)_B \leq 1$. Figure 2C shows a typical example of a Density of Ratios plot for the aggregate parameter sensitivities of a concentration variable. The aggregate parameter sensitivity of X_i, $S(X_i)$, is defined as the Euclidean norm of the vector whose coordinates are the sensitivities with respect to the individual parameters. [The numerical method makes it possible to use different functions of the parameter sensitivities to define an aggregate sensitivity, e.g. a weighted average of the sensitivities could be used when one knows the relative importance of the individual parameters in the model.] The ratio is defined as the aggregate parameter sensitivity in the reference model divided by the corresponding aggregate in the alternative model. Again, we see that the reference model has smaller sensitivities.

A comparison of the models on the basis of the 3rd Routh criterion for stability (Figure 2E) shows that the margin of stability is smaller for the reference model; however, the magnitude of the difference is very small with ratios always greater than 0.81. The modal class of this ratio is the one closest to 1 (defined has 0.995 < ratio <1), with more than 35% of the models. Thus, models with or without overall feedback inhibition have very similar stability boundaries. This indicates that local stability is probably not an important criterion in comparing the models, since they are very similar in this aspect. Figure 2G shows that the model with overall feedback inhibition is typically more responsive than the alternative model lacking this inhibition, although there are a few exceptions.

Statistical analysis of ratios

Figures 2B, D, F and 2H show the moving median plots (Alves and Savageau, 2000a) corresponding to the raw Density of Ratios plots in Figures 2A, C, E and 2G. As was mentioned previously, robust systems function more reproducibly. Figure 2B shows an example of a moving median plot for an individual parameter sensitivity. There are two regions in which there is no correlation between the sensitivity in the reference model and the ratio of sensitivities in the reference and alternative models. These two regions are separated by a region with a sharp change

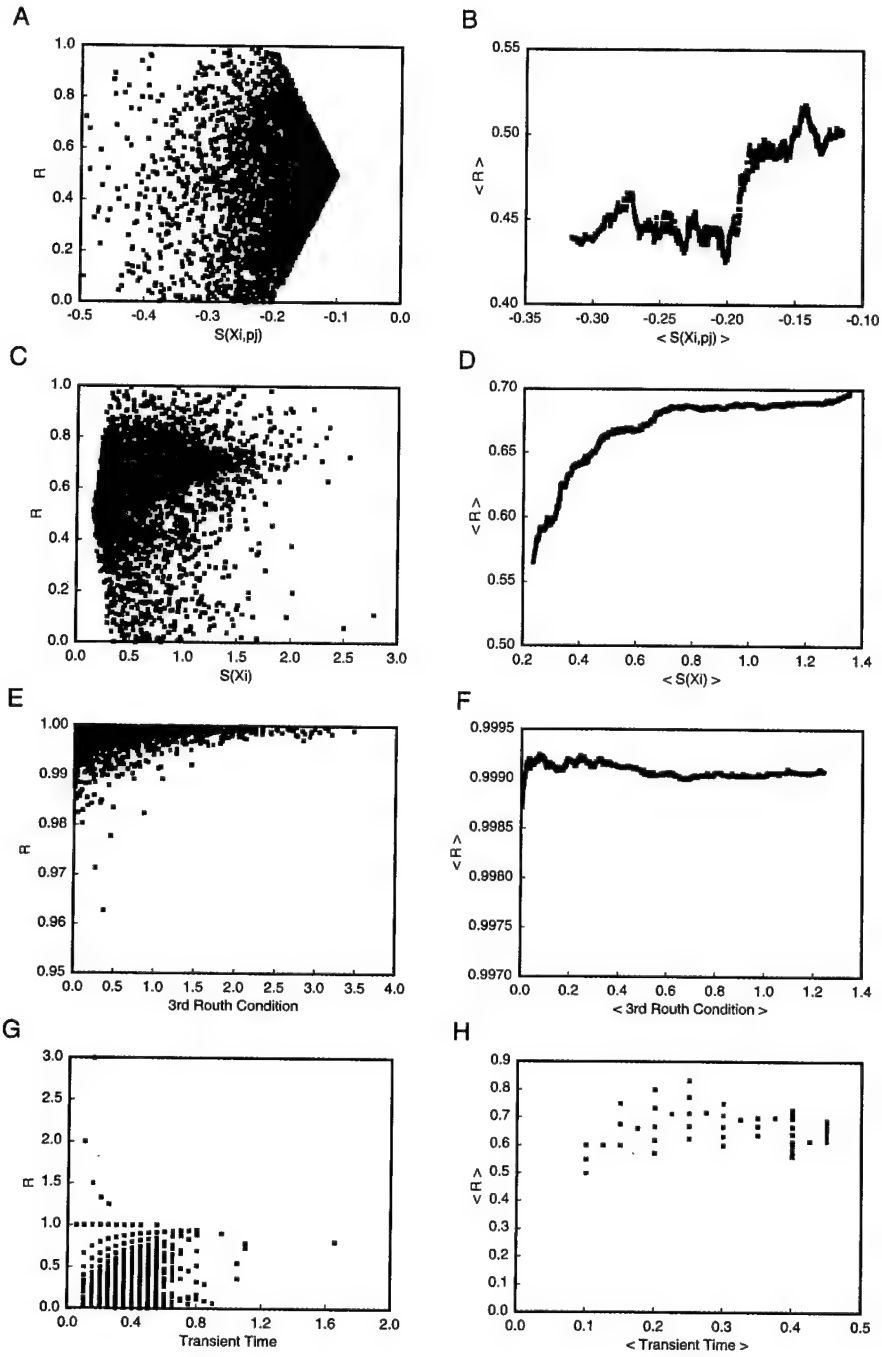


Fig. 2. Density of ratios plots for different systemic properties of the reference system vs the ratio of their values in the reference system to the corresponding value in the alternative system. Density of ratios plots for the primary data are shown in the left-hand panels (A, C, E and G) the corresponding moving median plots are shown in the right-hand panels (B, D, F and H). (A) and (B) a typical individual sensitivity, (C) and (D) a typical aggregate sensitivity (The aggregate sensitivity of any given metabolite is defined in this case as the Euclidean norm of a vector whose components are given by the sensitivity of the relevant metabolite to each of the parameters), (E) and (F) the Routh condition that differs between the reference and the alternative system, (G) and (H) the transient time.

in the average value of the ratio. For most other parameter sensitivities, the ratio changes less abruptly. The moving median plot for the aggregate parameter sensitivities (Figure 2D) shows that as the sensitivity increases (i.e. the robustness decreases) the ratio also tends to increase, until it reaches a limit median value. For highly robust models,

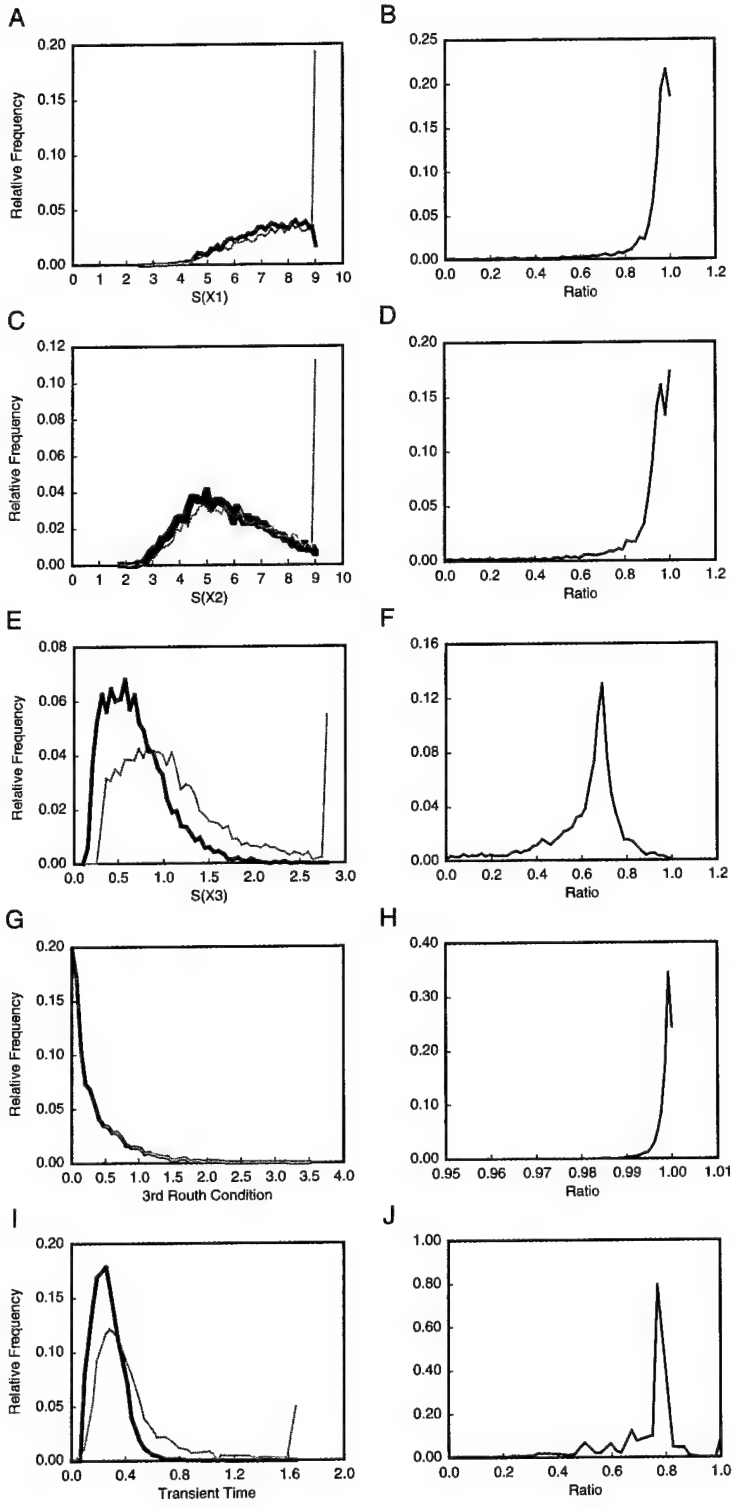


Fig. 3. Histograms comparing properties that differ between the reference system and the alternative system in Figure 1. In the left-hand panels (A, C, E, G and I) the histogram of the relevant property for the reference system is represented by a thick line whereas the same histogram for the alternative system is represented by a thin line. In the right-hand panels (B, D, F, H and J) the histogram for the ratio is represented with a thick line. (A) and (B) aggregate sensitivity of X_1, (C) and (D) aggregate sensitivity of X_2, (E) and (F) aggregate sensitivity of X_3, (G) and (H) 3rd Routh condition, (I) and (J) transient time.

the difference in robustness between the reference and the alternative model tends to be bigger. Thus, for models that are optimized with regard to robustness, on average, the reference model will be much more robust than the alternative model.

Figure 2F shows that the stability margin of the alternative model is always greater than that of the reference model, although on average the differences are insignificant. Hence, the stability margins are essentially the same.

As for the transient behavior of the models (Figure 2H), we can see that the responsiveness of the reference model is almost always better than or equal to that of the alternative model (more than 98% of all cases). Overall feedback inhibition has an important effect in making the model respond more quickly to perturbations in its state. Since there is no analytical expression for the transient behavior, the only way to obtain these comparative results is through the use of numerical methods.

A different way to observe that the parameter sensitivities are indeed larger in the alternative model, is by comparing histograms of corresponding sensitivities that differ between the reference and the alternative model and by plotting the histograms of the ratios directly (Figure 3). The alternative model clearly has more parameter sensitivities in the higher range of values. This approach also shows that the transient times are longer for the alternative model, whereas the difference in distributions for the stability margin is less notable. In each case, the histogram of ratios shows that the magnitudes are larger in the alternative model.

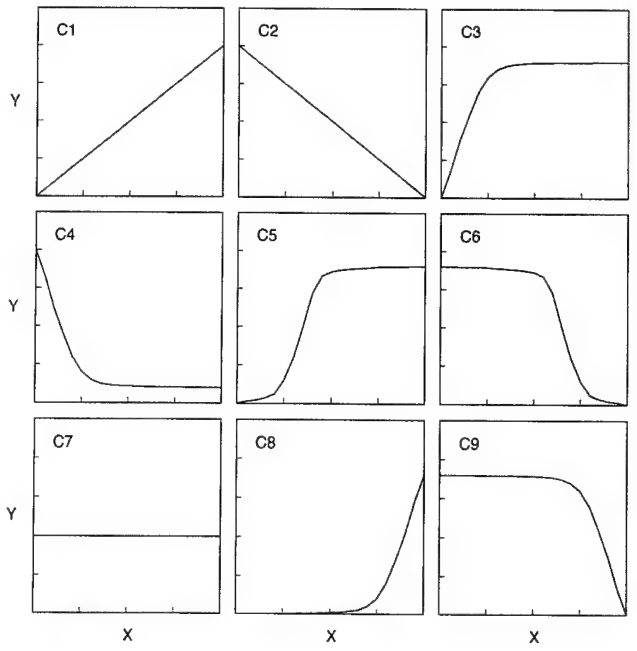


Fig. 4. Qualitatively different shapes for the correlation curves between different systemic properties. The correlation is determined by a plot of the moving median for one property versus the moving median for another constructed from an ensemble of systems selected on the basis of various behavioral classes (see Alves and Savageau, 2000b). The nine shapes (C1 through C9) include all the tabulated shapes found by examination of the actual graphs. These shapes are referenced in Table 2.

Correlations

The previous paper (Alves and Savageau, 2000b) has shown how different properties of the model represented in Figure 1A are correlated. Here we use the same technique to show how the differences between reference (Figure 1A) and alternative (Figure 1B) models are correlated with various steady-state properties. The differences we shall examine are the four analytical determined ratios shown in Table 1(A–D) plus the ratio of transient times that we determine numerically (E). For each of the five ratios we plot $\langle R \rangle_{Q_i}$ as a function of $\langle P \rangle$, where $\langle R \rangle_{Q_i}$ is the ith moving quantile of ratio R and $\langle P \rangle$ is the moving median of the steady-state property of interest. We present results for $i = 0.05$, $i = 0.5$ and $i = 0.95$. The moving window size used in the calculations is 500. The generic shapes of the correlation curves are shown in Figure 4, and the results of the correlation analysis are summarized with reference to these shapes in Table 2.

Each moving quantile curve for the same R, $\langle R \rangle_{Q_i}$, represents a contour that shows how a given quantile of R is correlated with a particular magnitude of interest. By building a contour plot with several different quantiles, we can empirically evaluate the quality of the predicted correlation, as well as obtain non-parametric confidence interval curves for the moving median.

An example of such a contour plot is presented in Figure 5 for the correlation between the different moving quantiles of the ratio B (from Table 1) and the 2nd Routh condition for local stability. The plot gives information about the dispersion of B as a function of the 2nd Routh criterion. This dispersion decreases as the value of the Routh criterion increases. At low values of the Routh criterion the 5% quantile of B is very close to -1 and the 95% quantile is very close to 1, whereas for high values of the stability margin the 5% quantile is about -0.6 and the 95% quantile is about 0.7. In plots involving other quantities, the dispersion may increase or remain unchanged as the quantity on the x-axis increases.

The second type of information one can extract from Figure 5 regards the quality of the predicted correlation between B and the 2nd Routh criterion. From the moving quantile plot involving $Q_{0.5}$ we determine that, on average, there is no correlation between B and the value for the 2nd Routh criterion. The other moving quantile curves show that, for high values of the stability margin, this absence of correlation is maintained for all moving quantiles. However, in the region of low values for the stability

Table 2. Correlation between the five critical ratios and various systemic properties

Systemic property	Critical ratios[a]														
	A			*B*			*C*			*D*			*E*		
	$Q_{0.05}$	$Q_{0.5}$	$Q_{0.95}$	$Q_{0.05}$	$Q_{0.5}$	$Q_{0.95}$	$Q_{0.05}$	$Q_{0.5}$	$Q_{0.95}$	$Q_{0.05}$	$Q_{0.5}$	$Q_{0.95}$	$Q_{0.05}$	$Q_{0.5}$	$Q_{0.95}$
X_1	C7[b]	C7	C7	C8	C7	C9	C9	C7	C7	C4	C4	C4	C7	C7	C7
X_2	C7	C7	C7	C2	C2	C2	C2	C7	C7	C4	C4	C4	C7	C7	C7
X_3	C8	C8	C8	C9	C2	C2	C9	C2	C2	C4	C4	C4	C7	C7	C7
V	C4	C4	C4	C4	C4	C4	C4	C4	C4	C4	C4	C4	C4	C4	C4
$L(X_1, X_0)$	C7	C7	C7	C7	C7	C7	C7	C7	C7	C7	C7	C7	C7	C7	C7
$L(X_2, X_0)$	C7	C7	C7	C7	C7	C7	C1	C7	C7	C7	C7	C2	C7	C7	C7
$L(X_3, X_0)$	C9	C9	C9	C7	C7	C7	C7	C7	C7	C2	C2	C2	C7	C7	C7
$L(V, X_0)$	C3	C3	C3	C2	C7	C1	C2	C7	C1	C7	C7	C7	C3	C3	C3
$L(X_1, X_4)$	C2	C2	C2	C9	C2	C4	C9	C2	C4	C2	C2	C2	C2	C2	C2
$L(X_2, X_4)$	C4	C4	C4	C9	C2	C4	C9	C2	C4	C4	C4	C4	C2	C2	C2
$L(X_3, X_4)$	C8	C7	C9	C9	C9	C9	C9	C9	C9	C9	C2	C2	C5	C7	C6
$L(V, X_4)$	C4	C2	C9	C3	C4	C4	C3	C4	C4	C4	C4	C2	C2	C2	C2
$S(X_1)$	C2	C2	C2	C3	C1	C1	C3	C1	C1	C3	C1	C1	C1	C7	C7
$S(X_2)$	C7	C7	C7	C1	C1	C7	C3	C1	C1	C1	C1	C1	C1	C1	C1
$S(X_3)$	C2	C7	C1	C1	C1	C1	C1	C1	C1	C7	C7	C7	C2	C1	C1
$S(V)$	C1	C7	C2	C2	C2	C2	C2	C2	C2	C2	C2	C2	C3	C3	C3
2nd Routh	C2	C2	C2	C3	C7	C4	C3	C7	C4	C3	C3	C3	C3	C7	C7
3rd Routh	C2	C2	C2	C3	C7	C4	C3	C7	C4	C3	C3	C3	C3	C3	C7
Transient time	C3	C3	C3	C2	C7	C1	C2	C7	C1	C3	C3	C3	C2	C7	C7

[a] The expressions for the critical ratios A, B, and C are given in the footnote of Table 1. The expression for ratio D is given following Equation (8) in the text. Ratio E is the ratio of transient times, which are determined numerically.

[b] The C values refer to the shape of the curves in Figure 5. We present the shape of the curves with a 90% confidence interval. For example, the correlation of ratio B with $L(X_1, X_4)$ has a shape C2 but its 90% boundaries show that this form can change (smoothly) between C9 and C4.

margin, a positive correlation between the stability margin and B starts to develop as the quantile of B decreases from $Q_{0.5}$ to $Q_{0.05}$. Symmetrically, a negative correlation develops as the quantile of B increases from $Q_{0.5}$ to $Q_{0.95}$. As Q_i tends to $Q_{0.0}$ or to $Q_{1.0}$ these correlations tend to be more pronounced. One interpretation is that, for low values of the stability margins, there is a larger uncertainty about the correlation between B and the stability margins.

Correlations among the four analytical determined ratios (A–C in Table 1 and D in equation (8)) plus the ratio of transient times that we determine numerically (E) are shown in Figure 6. It can be seen that the ratios A, D and E are directly correlated. This means that systems with high values (i.e. close to 1) for A will also, on average, have high values for D and E (i.e. close to 1). Similarly, the ratios B and C are directly correlated. On the other hand, the values of ratios B and C change from negatively to positively correlated with the other three ratios as the values of these other three ratios increases.

Summary

The numerical method reproduces the qualitative results that are obtained analytically as should be expected. Furthermore, the numerical comparison extends the analytical results by providing quantitative results. For example, overall feedback inhibition decreases the stability margins of the steady state, which was shown quantitatively to be on average a minimal effect, and increases the robustness of the system, which was shown quantitatively to be a highly significant effect. The numerical approach also provides a way to compare the temporal responsiveness of the alternative models following perturbations in the steady-state concentrations. For our model systems we found that overall feedback inhibition significantly decreases the response time of the reference system (with overall feedback), compared to that of the alternative system (without overall feedback). Finally, we determined how the different ratios are correlated with systemic properties and with parameters of interest, and we presented a way to determine the confidence one should place on these correlations. Thus, the numerical approach has significantly extended the scope of application beyond that of the analytical approach.

Discussion

The method of mathematically controlled comparison has been used since the early 1970s as a powerful tool to characterize alternative designs for several classes of biological systems. In each case, this comparative technique has provided insight into the natural selection of the various designs. The results obtained in some cases are independent of specific parameter values.

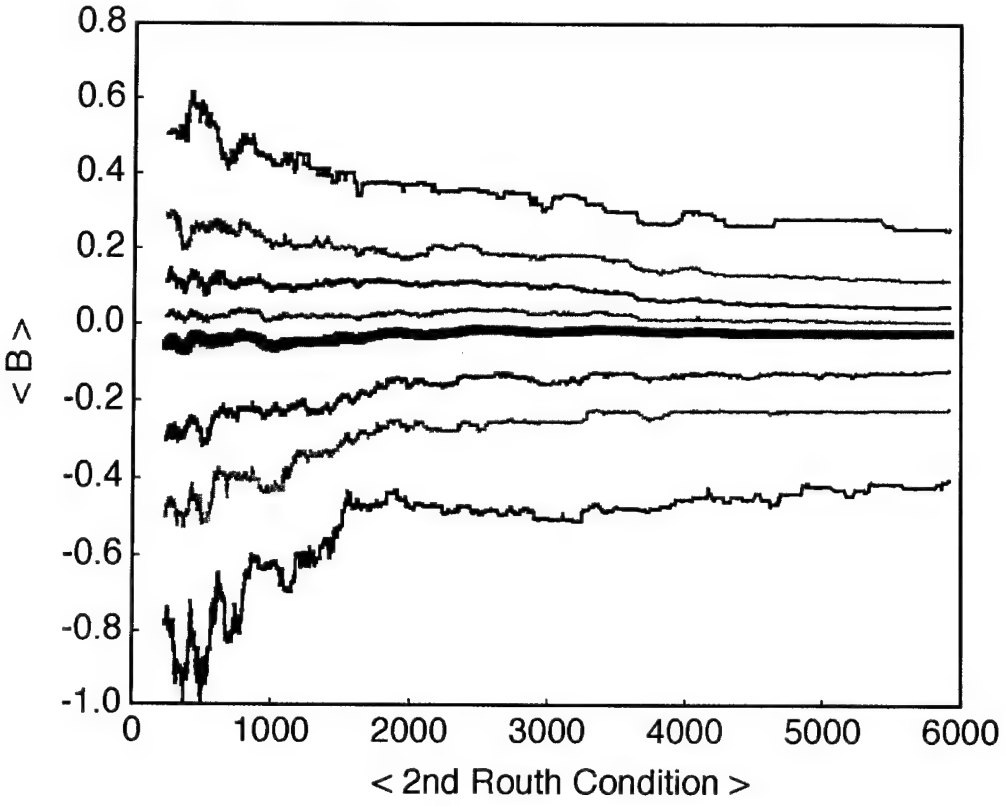


Fig. 5. Example of the use of moving quantiles to establish the level of confidence in the correlation between magnitudes of interest. The plot represents the correlation of different moving quantiles of the critical ratio B (defined in Table 1) as a function of the 2nd Routh condition. The thickest line represents the moving median of B. The first lines above and below the moving median represent quantiles 0.65 and 0.35 respectively. The progression of lines above and below represent quantiles that decrease and increase, respectively, by intervals of 0.10.

In many other cases, the results of the analysis will depend on the numerical values of the parameters. However, if the range of values for the parameters is known and their influence does not change abruptly over the range of interest, then random sampling can be used effectively to make numerical comparisons without exact knowledge of the parameter values. If one knows the distribution of values for each parameter, then one can generate random numbers with the appropriate distributions in order to obtain a large set of parameters and statistically study the differences between various designs. More often than not these distributions are unknown, because there are enormous numbers of components and interactions that need to be identified. In the absence of *a priori* knowledge about the distributions for the parameter values, we have generated random numbers with a uniform distribution and then refined the distributions by accepting vectors of parameter values only if they create a model that has the behavioral characteristics of interest e.g. in this paper, models with stable steady states; (see Alves and Savageau (2000b) for an analysis of models belonging to different behavioral classes).

Analytical comparisons in this paper demonstrate that the reference model is more robust and has smaller stability margins than the alternative model. They also show that the flux and concentrations in the reference model are less sensitive to changes in demand for end product. However, analytical comparison can not give us any qualitative information about the relative transient times of the two models, nor can it tell us anything quantitative about the magnitude of the differences in transient times between the two models.

The method of numerical comparison provides information about the alternative designs that could not have been obtained by exclusive use of analytical comparisons. It shows that the relative differences in parameter sensitivities and transient times between the reference and the alternative models are, on average, much larger than those between stability margins. This implies that differences in stability margins are not very relevant for the selection of overall feedback inhibition. Moreover, this approach shows that more than 99% of all reference models have faster transient responses than the corresponding alternative models. This reinforces the idea that overall feedback

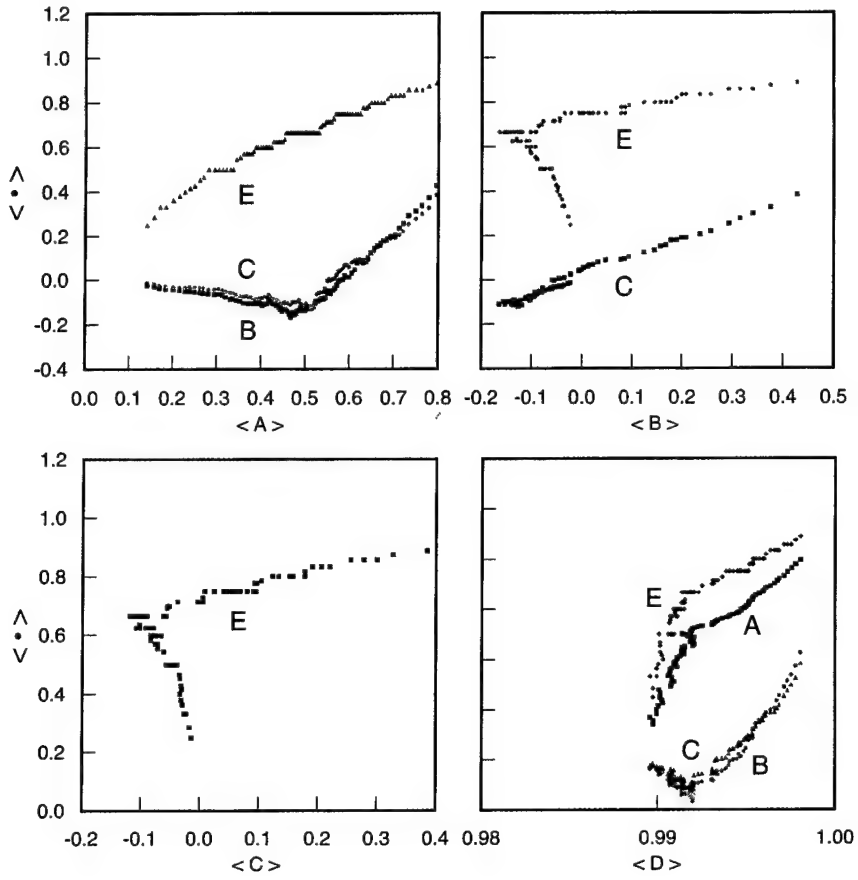


Fig. 6. Moving median plots that reveal the correlations among the five critical ratios (A–E) obtained while comparing the systems in Figure 1.

inhibition improves the function of these biosynthetic pathways.

With the method of numerical comparison we also show that, among the five critical ratios that characterize the alternative designs in Figure 1, there are two groups within which the ratios are directly correlated with each other. One group includes the ratios A, D and E; the other includes the ratios B and C. Members of the second group are negatively correlated with members of the first group at low values and positively correlated at high values.

The introduction of contour density of ratios plots (i.e. plots having different moving quantiles for the y-axis) provides a measure for the uncertainty in the correlations between ratios and systemic properties of interest. In most of the cases analyzed in this paper the correlation holds with a 90% confidence interval (i.e. the correlation is always positive, negative or null no matter what quantile is used, as can be seen in Table 2). However, in some cases, such as that in Figure 5, there is more uncertainty about the correlations. Although the nature of the correlations will be model and behavioral-class dependent, there are some properties of these contour plots that are general (see **Appendix**).

Thus, the method of numerical comparison presented in this paper allows one to quantify, and in some cases to eliminate, the uncertainties associated with the analytical approach to mathematically controlled comparison. This generalization allows one to obtain more information from the comparison. It also allows one to focus the comparison on systems that are considered most appropriate for each design, simply by selecting from randomly generated parameter values ensembles of parameter sets that give rise to systems that are considered appropriate with respect to properties of interest. This provides a means to make the comparison more significant.

Acknowledgements

This work was supported in part by a joint Ph.D. fellowship PRAXIS XXI / BD / 9803 / 96 granted by PRAXIS XXI through Programa Gulbenkian de Doutoramentos em Biologia e Medicina (R.A.), US Public Health Service

Grant RO1GM30054 from the National Institutes of Health (M.A.S.), and US Department of Defense Grant N00014-97-1-0364 from the Office of Naval Research (M.A.S.). We thank Armindo Salvador for critically reading early versions of this manuscript and making useful comments.

Appendix

Consider a contour density of ratios plot of property P_1 versus property P_2, such as the one presented in Figure 5. If $\langle P_1 \rangle_{Qk}$ represents the Q_k moving quantile curve for property P_1, $\langle P_1 \rangle_{Qm}$ represents the Q_m moving quantile curve for the same property, $k \leq m$, and the same window size is used to calculated the two curves, then contour density of ratios plots have the following generic properties:

(1) For any given value of $\langle P_2 \rangle$ on the x-axis, the curve $\langle P_1 \rangle_{Qk} \leq \langle P_1 \rangle_{Qm}$.

(2) The shape of the curve for $\langle P_1 \rangle_{Qr}$, with $k \leq r \leq m$, can only change progressively between the format of the curve $\langle P_1 \rangle_{Qk}$ and the format of the curve $\langle P_1 \rangle_{Qm}$.

The proof for the first property comes from the fact that, for the same value of $\langle P_2 \rangle$, $\langle P_1 \rangle_{Qk}$ and $\langle P_1 \rangle_{Qm}$ are different quantiles of the same sample. Thus, if $k \leq m$ then $Q_k \leq Q_m$.

The proof of the second property is also very simple. From Property 1 we know that $\langle P_1 \rangle_{Qk} \leq \langle P_1 \rangle_{Qm}$. Thus, $\langle P_1 \rangle_{Qk} \leq \langle P_1 \rangle_{Qr} \leq \langle P_1 \rangle_{Qm}$. At each value of $\langle P_2 \rangle$, the maximum number of different quantiles is W, which is the window size associated with the total sample of size S. Let $Q_{i/W}$ represent the quantile i/W, where i can vary from 1 to W. As $1/W \rightarrow 0$, such that the ratio $W/S \rightarrow$ constant, $\langle P_1 \rangle_{Qi/W} \rightarrow \langle P_1 \rangle_{Q(i+1)/W}$. Thus, since $\langle P_1 \rangle_{Q(i-1)/W} \leq \langle P_1 \rangle_{Qi/W} \leq \langle P_1 \rangle_{Q(i+1)/W}$, the shape of the curves will change progressively from $\langle P_1 \rangle_{Q0.0}$ to $\langle P_1 \rangle_{Q1.0}$.

References

Alves,R. and Savageau,M.A. (2000a) Comparing systemic properties of ensembles of biological networks by graphical and statistical methods. *Bioinformatics*, **16**, 527–533.

Alves,R. and Savageau,M.A. (2000b) Systemic properties of ensembles of metabolic networks: application of graphical and statistical methods to simple unbranched pathways. *Bioinformatics*, **16**, 534–547.

De Boer,R.J. and Hogeweg,P. (1989a) Stability of symmetric idiotypic networks—a critique of Hoffmann's analysis. *Bull. Math. Biol.*, **51**, 217–222.

De Boer,R.J. and Hogeweg,P. (1989b) Unreasonable implications of reasonableidiotypic network assumptions. *Bull. Math. Biol.*, **51**, 381–408.

De Boer,R.J. and Perelson,A.S. (1998) Target cell limited and immune control models of HIV infection: a comparison. *J. Theor. Biol.*, **190**, 201–214.

Dorf,R.C. (1992) *Modern Control Systems*. 6th edn., Addison-Wesley, Reading, MA.

Hamilton,J.D. (1994) *Time Series Analysis*. 2nd edn., Princeton University Press, New Jersey.

Hlavacek,W.S. and Savageau,M.A. (1996) Rules for coupled expression of regulator and effector genes in inducible circuits. *J. Mol. Biol.*, **255**, 121–139.

Hunding,A. (1974) Limit-cycles in enzyme-systems with nonlinear negative feedback. *Biophys. Struct. Mech.*, **1**, 47–54.

Irvine,D.H. (1991) The method of controlled mathematical comparisons. In Voit,E.O. (ed.), *Canonical Nonlinear Modeling: S-Systems Approach to Understanding Complexity*. Van Nostrand Reinhold, NY, pp. 90–109.

Irvine,D.H. and Savageau,M.A. (1985) Network regulation of the immune response: alternative control points for suppressor modulation of effector lymphocytes. *J. Immunol.*, **134**, 2100–2116.

Monod,J., Changeaux,J.P. and Jacob,F. (1963) Allosteric proteins and cellular control systems. *J. Mol. Biol.*, **6**, 306–329.

Savageau,M.A. (1969) Biochemical systems analysis II: the steady state solution for an n-pool system using a power law approximation. *J. Theor. Biol.*, **25**, 270–379.

Savageau,M.A. (1971) Concepts relating the behavior of biochemical systems to their underlying molecular properties. *Arch. Biochem. Biophys.*, **145**, 612–621.

Savageau,M.A. (1972) The behavior of intact biochemical control systems. *Curr. Top. Cell. Reg.*, **6**, 63–130.

Savageau,M.A. (1974) Optimal design of feedback control by inhibition: steady state considerations. *J. Mol. Evol.*, **4**, 139–156.

Savageau,M.A. (1976) *Biochemical Systems Analysis: A Study of Function and Design in Molecular Biology*. Addison-Wesley, Reading, Mass.

Savageau,M.A. (1996) Power-law formalism: a canonical nonlinear approach to modeling and analysis. In Lakshmikantham,V. (ed.), *World Congress of Nonlinear Analysts 92*. vol. 4, Walter de Gruyter Publishers, Berlin, pp. 3323–3334.

Savageau,M.A. and Jacknow,G. (1979) Feedforward inhibition in biosynthetic pathways: inhibition of the aminoacyl-tRNA synthetase by intermediaries of the pathway. *J. Theor. Biol.*, **77**, 405–425.

Shiraishi,F. and Savageau,M.A. (1992) The tricarboxylic acid cycle in *Dictyostelium discoideum* II. Evaluation of model consistency and robustness. *J. Biol. Chem.*, **267**, 22919–22925.

Umbarger,H.E. (1956) Evidence for a negative-feedback mechanism in the biosynthesis of isoleucine. *Science*, **123**, 848.

Wherry,R.J. (1984) *Contributions to Correlational Analysis*. Academic Press, Orlando.

Wolfram,S. (1997) *Mathematica™: A System for Doing Mathematics by Computer*. Addison-Wesley, Menlo Park, CA.

Yates,R.A. and Pardee,A.B. (1956) Control of pyrimidine biosynthesis in *E.coli* by a feedback mechanism. *J. Biol. Chem.*, **221**, 757–770.

Design principles for elementary gene circuits: Elements, methods, and examples

Michael A. Savageau[a)]
Department of Microbiology and Immunology, University of Michigan Medical School, 5641 Medical Science Building II, Ann Arbor, Michigan 48109-0620

(Received 25 July 2000; accepted for publication 19 December 2000)

The control of gene expression involves complex circuits that exhibit enormous variation in design. For years the most convenient explanation for these variations was historical accident. According to this view, evolution is a haphazard process in which many different designs are generated by chance; there are many ways to accomplish the same thing, and so no further meaning can be attached to such different but equivalent designs. In recent years a more satisfying explanation based on design principles has been found for at least certain aspects of gene circuitry. By design principle we mean a rule that characterizes some biological feature exhibited by a class of systems such that discovery of the rule allows one not only to understand known instances but also to predict new instances within the class. The central importance of gene regulation in modern molecular biology provides strong motivation to search for more of these underlying design principles. The search is in its infancy and there are undoubtedly many design principles that remain to be discovered. The focus of this three-part review will be the class of elementary gene circuits in bacteria. The first part reviews several elements of design that enter into the characterization of elementary gene circuits in prokaryotic organisms. Each of these elements exhibits a variety of realizations whose meaning is generally unclear. The second part reviews mathematical methods used to represent, analyze, and compare alternative designs. Emphasis is placed on particular methods that have been used successfully to identify design principles for elementary gene circuits. The third part reviews four design principles that make specific predictions regarding (1) two alternative modes of gene control, (2) three patterns of coupling gene expression in elementary circuits, (3) two types of switches in inducible gene circuits, and (4) the realizability of alternative gene circuits and their response to phased environmental cues. In each case, the predictions are supported by experimental evidence. These results are important for understanding the function, design, and evolution of elementary gene circuits. © *2001 American Institute of Physics.* [DOI: 10.1063/1.1349892]

Gene circuits sense their environmental context and orchestrate the expression of a set of genes to produce appropriate patterns of cellular response. The importance of this role has made the experimental study of gene regulation central to nearly all areas of modern molecular biology. The fruits of several decades of intensive investigation have been the discovery of a plethora of both molecular mechanisms and circuitry by which these are interconnected. Despite this impressive progress we are at a loss to understand the integrated behavior of most gene circuits. Our understanding is still fragmentary and descriptive; we know little of the underlying design principles. Several elements of design, each exhibiting a variety of realizations, have been identified among elementary gene circuits in prokaryotic organisms. The use of well-controlled mathematical comparisons has revealed design principles that appear to govern the realization of these elements. These design principles, which make specific predictions supported by experimental data, are important for understanding the normal function of gene circuits; they also are potentially important for developing judicious methods to redirect normal expression for biotechnological purposes or to correct pathological expression for therapeutic purposes.

I. INTRODUCTION

The gene circuitry of an organism connects its gene set (genome) to its patterns of phenotypic expression. The genotype is determined by the information encoded in the DNA sequence, the phenotype is determined by the context-dependent expression of the genome, and the circuitry interprets the context and orchestrates the patterns of expression. From this perspective it is clear that gene circuitry is at the heart of modern molecular biology. However, the situation is considerably more complex than this simple overview would suggest. Experimental studies of specific gene systems by molecular biologists have revealed an immense variety of molecular mechanisms that are combined into complex gene circuits, and the patterns of gene expression observed in response to environmental and developmental signals are equally diverse.

The enormous variety of mechanisms and circuitry raises questions about the bases for this diversity. Are these variations in design the result of historical accident or have they been selected for specific functional reasons? Are there design principles that can be discovered? By design principle

a)Electronic mail: savageau@umich.edu

 © 2001 American Institute of Physics

Downloaded 20 Mar 2001 to 141.214.51.241. Redistribution subject to AIP copyright, see http://ojps.aip.org/chaos/chocr.jsp

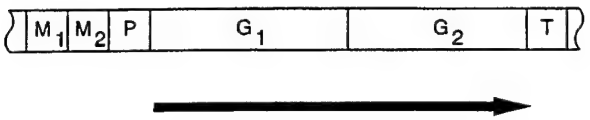


FIG. 1. Schematic diagram of a bacterial transcription unit. The structure of the unit consists of two genes (G_1 and G_2), bounded by a promoter sequence (P) and a terminator sequence (T), and preceded by upstream modulator sites (M_1 and M_2) that bind regulators capable of altering transcription initiation. The solid arrow represents the mRNA transcript.

we mean a rule that characterizes some biological feature exhibited by a class of systems such that discovery of the rule allows one not only to understand known instances but also to predict new instances within the class. For many years, most molecular biologists assumed that accident played the dominant role, and the search for rules received little attention. More recently, simple rules have been identified for a few variations in design. Accident and rule both have a role in evolution and their interplay has become clearer in these well-studied cases. This area of investigation is in its infancy and many such questions remain unanswered.

This review article addresses the search for design principles among elementary gene circuits. It reviews first several elements of design for gene circuits, then mathematical methods used to study variations in design, and finally examples of design principles that have been discovered for elementary gene circuits in prokaryotes.

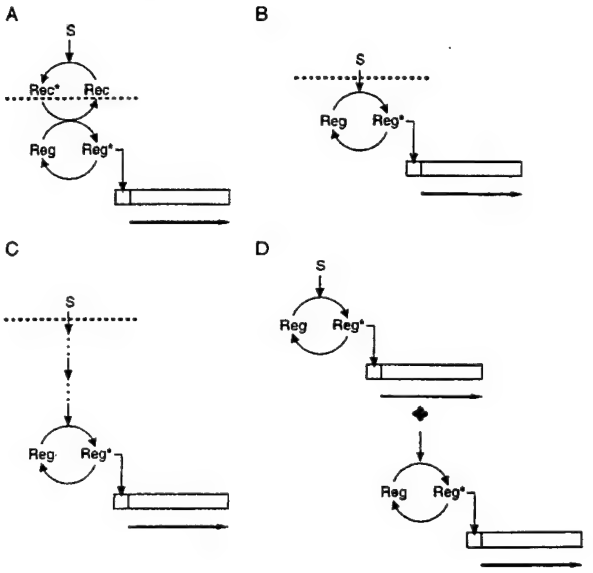


FIG. 2. Input signals for transcription units can arise either from the extracellular environment or from within the cell. S is a stimulus, Rec and Rec* are the inactive and active forms of the receptor, and Reg and Reg* are the inactive and active forms of the regulator. (a) Signal transduction from the extracellular environment to an intracellular transcription unit via a two-component system. (B) The extracellular signal molecule is transported into the cell where it interacts directly with the regulator of a transcription unit. (c) The signal molecule is transported into the cell where it is transformed via a metabolic pathway to produce a product that interacts with the regulator of a transcription unit. (d) The output signal from one transcription unit is the input signal to another transcription unit within the cell.

II. ELEMENTS OF DESIGN AND THE NEED FOR DESIGN PRINCIPLES

The behavior of an intact biological system can seldom be related directly to its underlying genome. There are several different levels of hierarchical organization that intervene between the genotype and the phenotype. These levels are linked by gene circuits that can be characterized in terms of the following elements of design: transcription unit, input signaling, mode of control, logic unit, expression cascade, and connectivity. Each of these elements exhibits a variety of realizations whose basis is poorly understood.

A. Transcription unit

A landmark in our understanding of gene circuitry was the discovery by Jacob and Monod of the operon,[1] the simplest of transcription units. This unit of sequence organization consists of a set of coordinately regulated structural genes (e.g., G_1 and G_2 in Fig. 1) that encode proteins, an up-stream promoter site (P) at which transcription of the genes is initiated, and a down-stream terminator site (T) at which transcription ceases. Modulator sites (e.g., M_1 and M_2 in Fig. 1) associated with the promoter bind regulatory proteins that influence the rate of transcription initiation (operator sites bind regressors that down-regulate high-level promoters, or initiator sites bind activators that up-regulate low-level promoters).

Transcription units are the principal feature around which gene circuits are organized. On the input side, signals in the extracellular (or intracellular) environment are detected by binding to specific receptor molecules, which propagate the signal to specific regulatory molecules in a process called transduction, although in many cases the regulator molecules are also the receptor molecules. Regulator molecules in turn bind to the modulator sites of transcription units in one of two alternative modes, and the signals are combined in a logic unit to determine the rate of transcription. On the output side, transcription initiates an expression cascade that yields one or many mRNA products, one or many protein products, and possibly one or many products of enzymatic activity. Thus, the transcription unit emits a fan-out of signals, which are then connected in a diverse fashion to the receptors of other transcription units to complete the interlocking gene circuitry.

B. Input signaling

The input signals for transcription units can arise either from the external environment or from within the cell. When signals originate in the extracellular environment, they often involve binding of signal molecules to specific receptors in the cellular membrane [Fig. 2(a)]. In bacteria, alterations in the membrane-bound receptor are communicated directly to regulator proteins via short signal transduction pathways called "two-component systems."[2] In other cases, signal molecules in the environment are transported across the membrane [Fig. 2(b)], and in some cases are subsequently modified metabolically [Fig. 2(c)], to become signal molecules that bind directly to regulator proteins (in these cases

Downloaded 20 Mar 2001 to 141.214.51.241. Redistribution subject to AIP copyright, see http://ojps.aip.org/chaos/chocr.jsp

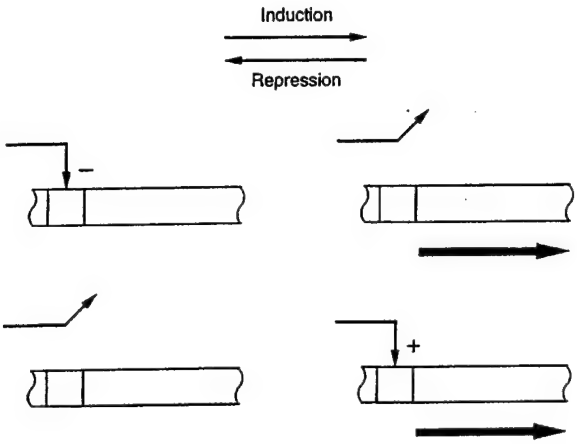
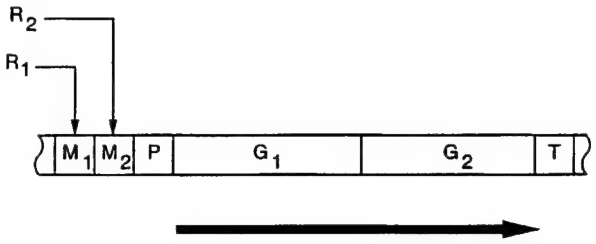


FIG. 3. Alternative modes of gene control. The top panels illustrate the negative mode of control in which the bias for expression is ON in the absence of the regulator, and regulation is achieved by modulating the effectiveness of a negative element. The bottom panels illustrate the positive mode of control in which the bias for expression is OFF in the absence of the regulator, and regulation is achieved by modulating the effectiveness of a positive element. The solid arrow represents the mRNA transcript. In each case, induction by the addition of a specific inducer causes the state of the system to shift from the left to the right, whereas repression by the addition of a specific co-repressor causes the state of the system to shift from right to left.



Mode		Presence		Expression	
R_1	R_2	R_1	R_2	AND	OR
+	–	Yes	Yes	OFF	ON
		Yes	No	**ON**	ON
		No	Yes	OFF	**OFF**
		No	No	OFF	ON

FIG. 4. Logic unit with two inputs. The transcription unit is described in Fig. 1, the regulator R_1 interacts with the modulator site M_1 via the positive mode, the regulator R_2 interacts with the modulator sites M_2 via the negative mode, and the signals are combined by a simple logical function. The logic table is provided for the logical AND and logical OR functions.

the receptor and regulator are one and the same molecule). When signals arise from other transcription units within the cell, the regulator can be the direct output signal from such a transcription unit [Fig. 2(d)]. It can also be the terminus of a signal transduction pathway in which the upstream signal is the output from such a transcription unit. Thus, the input signals for transcription units are ultimately the regulators, whether signals are received from the extracellular or intracellular environment. The regulators in most cases are protein molecules, although this function can be preformed in some cases by other types of molecules such as anti-sense RNA.

C. Mode of control

Regulators exert their control over gene expression by acting in one of two different modes (Fig. 3).[3] In the positive mode, they stimulate expression of an otherwise quiescent gene, and induction of gene expression is achieved by supplying the functional form of the regulator. In the negative mode, regulators block expression of an otherwise active gene, and induction of gene expression is achieved by removing the functional form of the regulator. Each of these two designs (positive or negative) requires the transcription unit to have the appropriate modulator site (initiator type or operator type) and promoter function (low level or high level).

Variations in the level of the functional form of the regulator can be achieved in different ways. Regulator molecules can have a constant or constitutive level of expression. In this case, the functional form of the regulator is created or destroyed by molecular alterations associated with the binding of specific ligands (inducers or co-regressors). In other cases, the regulator is always in the functional form, and its level of expression varies as the result of changes in its rate of synthesis or degradation. These different ways of realizing variations in the functional form of the regulator are found for both positive and negative modes of control.

D. Logic unit

The control regions associated with transcription units may be considered the logic unit where input signals from various regulators are integrated to govern the rate of transcription initiation. There are two lines of evidence suggesting that most transcription units in bacteria have only a few regulatory inputs. First, the early computational studies of Stuart Kauffman using abstract random Boolean networks suggested that two or three inputs per transcription unit were optimal.[4] If the number of inputs was fewer on average, the behavior of the network was too fixed; whereas if the number was greater on average, the behavior was too chaotic. The optimal behavior associated with a few inputs often is described as "operating at the edge of chaos."[5] Second, with the arrival of the genomic era and the sequencing of the complete genome for a number of bacteria, there is now experimental evidence regarding the distribution of inputs per transcription unit. The sequence for *Escherichia coli*[6] has shown that the number of modulator sites located near the promoters of transcription units is on average approximately two to three.[7] The large majority have two and a few have as many as five.

A simple logic unit is illustrated in Fig. 4 for the case with two inputs. This example includes the classical lactose

Downloaded 20 Mar 2001 to 141.214.51.241. Redistribution subject to AIP copyright, see http://ojps.aip.org/chaos/chocr.jsp

(*lac*) operon of *E. coli*, which has a positive and a negative regulator; the AND function is the logical operator by which these signals are combined.[8] The logic units of eukaryotes can be considerably more complex.[9]

E. Expression cascades

Expression cascades produce the output signals from transcription units. They typically reflect the flow of information from DNA to RNA to protein to metabolites, which has been called the "Central Dogma" of molecular biology. The initial output of a transcription unit is an mRNA molecule that has a sequence complementary to the transcribed DNA strand. The mRNA in turn is translated to produce the encoded protein product. The protein product in many instances is an enzyme, which in turn catalyzes a specific reaction to produce a particular metabolic product. This in skeletal form is the expression cascade that is initiated by signals affecting a transcription unit (Fig. 5).

There are many variations on this theme. There can be additional stages in such cascades and each of the stages is a potential target for regulation. For example, the cascade might include posttranscriptional or posttranslational stages in which products are processed before the next stage in the cascade. The cascade can also include a stage in which a RNA template is used to transcribe a complementary DNA copy, as is the case with retroviruses and retrotransposons.

There can be multiple products produced at each stage of such cascades. For example, several different mRNA molecules can arise from the same transcription unit by regulation of transcription termination. Several different proteins can be synthesized from the same mRNA and this is often the case in bacteria. Several metabolic products can be produced by a given multifunctional enzyme, depending upon its modular composition. Thus, transcription units can be considered to emit a fan of output signals.

F. Connectivity

The connectivity of gene circuits, defined as the manner in which the outputs of transcription units are connected to the inputs of other transcription units, varies enormously. The evidence for *E. coli* suggests a fairly narrow distribution of input connections with a mean of two to three, whereas the distribution of output connections has a wider distribution with some transcription units having as many as 50 output connections. A large number of the connections involve

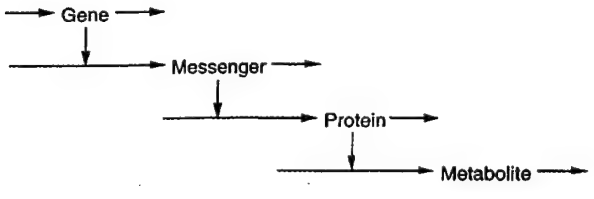


FIG. 5. Expression cascade that propagates signals in three stages from DNA to mRNA to enzymes to small molecular weight signaling molecules. Additional stages are possible, and each stage can give rise to multiple output signals.

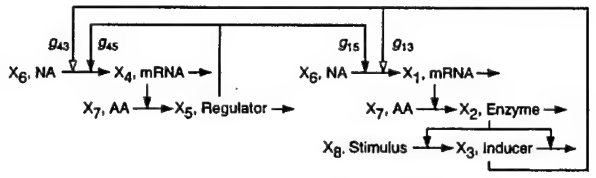


FIG. 6. Connectivity by which expression cascades become coupled. Elementary circuit consisting of a regulator cascade on the left and an effector cascade on the right. The protein product that is the output of the left cascade is a regulator of both transcription units, and the metabolic intermediate that is an output of the right cascade is an inducer that modulates the effectiveness of the regulator at each transcription unit.

regulator proteins modulating expression of the transcription unit in which they are encoded, a form of regulation termed autogenous.[10] Another common form of connection involves the coupling of expression cascades for an effector function and for its associated regulator.[11] Such couplings are called elementary gene circuits and an example is represented schematically in Fig. 6.

Connectivity provides a way of coordinating the expression of related functions in the cell.[12] The operon, a transcription unit consisting of several structural genes that are transcribed as a single polycistronic mRNA, provides one way of coordinating the expression of several genes. Another way is to have each gene in a separate transcription unit and have all the transcription units connected to the same regulatory input signal. Such a set of coordinately regulated transcription units is known as a regulon. Other, and more flexible, ways also exist. For example, when signals from several regulators are assembled in a combinatorial fashion to govern a collection of transcription units, each with its own logic unit, diverse patterns of gene expression can be orchestrated in response to a variety of environmental contexts.

III. METHODS FOR COMPARING DESIGNS TO REVEAL DESIGN PRINCIPLES

Several different approaches have been used to analyze and compare gene circuits, and each has contributed in different ways to our understanding. Here I need only mention three of the approaches that have been dealt with in greater detail elsewhere.

A. Types of models

Simplified models based on random Boolean networks have been used to explore properties that are likely to be present with high probability regardless of mechanistic details or evolutionary history. These tend to be discrete/deterministic models that permit efficient computational exploration of large populations of networks, which then permit statistical conclusions to be drawn. The work of Kauffman provides an example of this approach.[4] The elements of design emphasized in this approach are the input logic units and the connectivity, and properties of the network are examined as a function of network size.

Detailed mechanistic models have been used to test our understanding of particular gene circuits. The goal is to represent the detailed behavior as faithfully as possible. A mixture of discrete/continuous/deterministic/stochastic model el-

Downloaded 20 Mar 2001 to 141.214.51.241. Redistribution subject to AIP copyright, see http://ojps.aip.org/chaos/chocr.jsp

ements might be used, depending upon the particular circuit. These are computationally intensive and require numerical values for the parameters, and detailed quantitative comparisons with experimental data are important means of validation. The work of Arkin and colleagues illustrates this approach.[13] The elements of design emphasized in this approach are all those that manifest themselves in the particular circuit being modeled.

Generic models for specific classes of circuits have been used to identify design principles for each class. The aim of these models is to capture qualitative features of behavior that hold regardless of the specific values for the parameters and hence that are applicable to the entire class being characterized. These tend to be continuous/deterministic models with a regular formal structure that facilitates analytical and numerical comparisons. Examples of this approach will be reviewed below in Sec. IV. The elements of design that tend to be emphasized in this approach are expression cascades, modes of control, input logic units, and connectivity.

B. A comparative approach to the study of design

The elucidation of design principles for a class of circuits requires a formalism to represent alternative designs, methods of analysis capable of predicting behavior, and methods for making well-controlled comparisons.

1. Canonical nonlinear representation

The power-law formalism combines nonlinear elements having a very specific structure (products of power laws) with a linear operator (differentiation) to form a set of ordinary differential equations, which are capable of representing any suitably differentiable nonlinear function. This makes it an appropriate formalism for representing alternative designs.

The elements of the power-law formalism are nonlinear functions consisting of simple products of power-law functions of the state variables[14]

$$v_i(\mathbf{X}) = \alpha_i X_1^{g_{i1}} X_2^{g_{i2}} X_3^{g_{i3}} \cdots X_n^{g_{in}}. \tag{1}$$

The two types of parameters in this formalism are referred to as *multiplicative parameters* (α_i) and *exponential parameters* (g_{ij}). They also are referred to as *rate-constant parameters* and *kinetic-order parameters*, since these are accepted terms in the context of chemical and biochemical kinetics. The multiplicative parameters are non-negative real, the exponential parameters are real, and the state variables are positive real.

Although the nonlinear behavior exhibited by these nonlinear elements is fairly impressive, it does not represent the full spectrum of nonlinear behavior that is characteristic of the power-law formalism. When these nonlinear elements are combined with the differential operator to form a set of ordinary differential equations they are capable of representing any suitably differentiable nonlinear function. The two most common representations within the power-law formalism are generalized-mass-action (GMA) systems

$$dX_i/dt = \sum_{k=1}^{r} \alpha_{ik} \prod_{j=1}^{n+m} X_j^{g_{ijk}} - \sum_{k=1}^{r} \beta_{ik} \prod_{j=1}^{n+m} X_j^{h_{ijk}}, \quad i=1,\ldots,n, \tag{2}$$

and synergistic (S) systems

$$dX_i/dt = \alpha_i \prod_{j=1}^{n+m} X_j^{g_{ij}} - \beta_i \prod_{j=1}^{n+m} X_j^{h_{ij}}, \quad i=1,\ldots,n. \tag{3}$$

The derivatives of the state variables with respect to time t are given by dX_i/dt. The α and g parameters are defined as in Eq. (1) and are used to characterize the positive terms in Eqs. (2) and (3), whereas the β and h parameters are similarly defined and are used to characterize the negative terms. There are in general n dependent variables, m independent variables, and a maximum of r terms of a given sign. The resulting power-law formalism can be considered a canonical nonlinear representation from at least three different perspectives: fundamental, recast, and local.[15]

As the natural representation of the elements postulated to be fundamental in a variety of fields, the power-law formalism can be considered a canonical nonlinear representation. There are a number of representations that are considered fundamental descriptions of the basic entities in various fields. Four such representations that are extensively used in chemistry, population biology, and physiology are mass-action, Volterra–Lotka, Michaelis–Menten, and linear representations. These are, in fact, special cases of the GMA-system representation,[15] which, as noted earlier, is one of the two most common representations within the general framework of the power-law formalism. Although, the power-law formalism can be considered a fundamental representation of chemical kinetic events, this is not the most useful level of representation for comparing gene circuits because it is much too detailed and values for many of the elementary parameters will not be available. Nor does the structure of the GMA equations lend itself to general symbolic analysis.

As a recast description, the power-law formalism can be considered a canonical nonlinear representation in nearly every case of physical interest. This is because any nonlinear function or set of differential equations that is a composite of elementary functions can be transformed exactly into the power-law formalism through a procedure called *recasting*.[16] This is a well-defined procedure for generating a globally accurate representation that is functionally equivalent to the original representation. In this procedure one trades fewer equations with more complex and varied forms of nonlinearity for more equations with simpler and more regular nonlinear forms. Although the power-law formalism in the context of recasting has important uses and allows for efficient numerical solution of differential equations, this again is not the most useful level of representation for comparing alternative designs for gene circuits because it does not lend itself to general systematic analysis.

As a local description, the power-law formalism can be considered a canonical nonlinear representation that is typically accurate over a wider range of variation than the corresponding linear representation. The state variables of a system can nearly always be defined as positive quantities. Therefore, functions of the state variables can be represented

Downloaded 20 Mar 2001 to 141.214.51.241. Redistribution subject to AIP copyright, see http://ojps.aip.org/chaos/chocr.jsp

equivalently in a logarithmic space—i.e., a space in which the logarithm of the function is a function of the logarithms of the state variables. This means that a Taylor series in logarithmic space can also be used as a canonical representation of the function. If the variables make only small excursions about their nominal operating values, then this series can be truncated at the linear terms, transformed back into Cartesian coordinates, and expressed in the power-law formalism. Thus, Taylor's theorem gives a rigorous justification for the local power-law formalism and specific error bounds within which it will provide an accurate representation.

A rigorous and systematic analysis of the second-order contributions to the local power-law representation has been developed by Salvador.[17,18] This analysis provides a valuable approach for making rational choices concerning model reduction. By determining the second-order terms in the power-law approximation of a more complex model one can determine those parts of the model that are accurately represented by the first-order terms. These parts of the model can be safely represented by the local representation; those parts that would not be represented with sufficient accuracy can then be dealt with in a variety of ways, including a more fundamental model or a recast model, either of which would leave the resulting model within the power-law representation.

The local S-system representation within the power-law formalism has proved to be more fruitful than the local GMA-system representation because of its accuracy and structure. It is typically more accurate because it allows for cancellation of systematic errors.[19,20] It has a more desirable structure from the standpoint of general symbolic analysis: there is an analytical condition for the existence of a steady state, an analytical solution for the steady state, and an analytical condition that is necessary for the local stability of the steady state. The regular structure and tractability of the S-system representation is an advantage in systematic approaches for inferring the structure of gene networks from global expression data.[21]

The S-system representation, like the linear and Volterra–Lotka representations, exhibits the same structure at different hierarchical levels of organization.[22] We call this the *telescopic* property of the formalism. Only a few formalisms are known to exhibit this property. A canonical formalism that provides a consistent representation across various levels of hierarchical organization in space and time has a number of advantages. For example, consider a system described by a set of S-system equations with n dependent variables. Now suppose that the variables of the system form a temporal hierarchy such that k of them determine the temporal behavior of the system. The $n-k$ "fast" variables are further assumed to approach a quasi-steady state in which they are now related to the k temporally dominant variables by power-law equations. When these relationships are substituted into the differential equations for the temporally dominant variables, a new set of differential equations with k dependent variables is the result. This reduced set is also an S-system; that is, the temporally dominant subsystem is represented within the same power-law formalism. Thus, the same methods of analysis can be applied at each hierarchical level.

Power-law expressions are found at all hierarchical levels of organization from the molecular level of elementary chemical reactions to the organismal level of growth and allometric morphogenesis.[15] This recurrence of the power law at different levels of organization is reminiscent of fractal phenomena, which exhibit the same behavior regardless of scale.[23] In the case of fractal phenomena, it has been shown that this self-similar property is intimately associated with the power-law expression.[24] Hence, it is not surprising that the power-law formalism should provide a canonical representation with telescopic properties appropriate for the characterization of complex nonlinear systems.

Finally, piecewise power-law representations provide a logical extension of the local power-law representation. The piecewise *linear* representation has long been used in the temporal analysis of electronic circuits.[25] It simplifies the analysis, converting an intractable nonlinear system of equations into a series of simple linear systems of equations whose behavior, when pieced together, is capable of closely approximating that of the original system. A different use of an analogous piecewise representation was developed by Bode to simplify the interpretation of complex rational functions that characterize the frequency response of electronic circuits.[26] This type of Bode analysis was adapted for interpretation of the rational functions traditionally used to represent biochemical rate laws[27] and then developed more fully into a systematic power-law formalism for the local representation of biochemical systems consisting of many enzymatic reactions.[15] In analogy with traditional piecewise linear analysis, a piecewise power-law representation has been developed and used to analyze models of gene circuitry (see Sec. IV C). This form of representation greatly simplifies the analysis; it also captures the essential nonlinear behavior more directly and with fewer segments than would a piecewise linear representation.

2. Methods of analysis

The regular, systematic structure of the power-law formalism implies that methods developed to solve efficiently equations having this form will be applicable to a wide class of phenomena. This provides a powerful stimulus to search for such methods. The potential of the power-law formalism in this regard has yet to be fully exploited. The following are some examples of generic methods that have been developed for analysis within the framework of the power-law formalism.

The simplicity of the local S-system representation has led to the most extensive development of theory, methodology, and applications within the power-law formalism.[28] Indeed, as discussed in Sec. III B 1, the local S-system representation allows the derivation of important systemic properties that would be difficult, if not impossible, to deduce by other means. These advances have occurred because it was recognized from the beginning that the steady-state analysis of S-systems reduces to conventional linear analysis in a logarithmic space. Hence, one was able to exploit the powerful methods already developed for linear systems. For

Downloaded 20 Mar 2001 to 141.214.51.241. Redistribution subject to AIP copyright, see http://ojps.aip.org/chaos/chocr.jsp

example, S-systems have an explicit analytical solution for the steady state.[14,27] The condition for the existence of such a steady state reduces to the evaluation of a simple determinant involving the exponential parameters of the S-system. Local stability is determined by two critical conditions, one involving only the exponential parameters and the other involving these as well as the multiplicative parameters. Steady-state (logarithmic) gain matrices provide a complete network analysis of the signals that propagate through the system. Similarly, steady-state sensitivity matrices provide a complete sensitivity analysis of the parameters that define the system and its robustness. The linear structure also permits the use of well-developed optimization theory such as the simplex method.[29]

Analytical solutions for the local dynamic behavior are available, including eigenvalue analysis for characterization of the relaxation times.[30] The regular structure also allows the conditions for Hopf bifurcation to be expressed as a simple formula involving the exponential parameters.[31] However, S-systems are ultimately nonlinear systems and so there is no analytical solution for dynamic behavior outside the range of accurate linear representation, which is more restrictive than the range of accurate power-law representation. Determination of the local dynamic behavior within this larger range, and the determination of global dynamic behavior, requires numerical methods.

An example of what can be done along these lines is the efficient algorithm developed for solving differential equations represented in the canonical power-law formalism.[32] This algorithm, when combined with recasting,[16] can be used to obtain solutions for rather arbitrary nonlinear differential equations. More significantly, this canonical approach has been shown to yield solutions in shorter time and with greater accuracy, reliability, and predictability than is typically possible with other methods. This algorithm can be applied to other canonical formalisms as well as to all representations within the power-law formalism. This algorithm has been implemented in a user-friendly program call PLAS (Power-Law Analysis and Simulation), which is available on the web (http://correio.cc.fc.ul.pt/~aenf/plas.html).

Another example is an algorithm based on the S-system representation that finds multiple roots of nonlinear algebraic equations.[33,34] Recasting allows one to express rather general nonlinear equations in the GMA-system representation within power-law formalism. The steady states of the GMA-system, which correspond to the roots of the original algebraic equation, cannot be obtained analytically. However, these power-law equations can be solved iteratively using a local S-system representation, which amounts to a Newton method in logarithmic space. Each step makes use of the analytical solution that is available with the S-system representation (see earlier in this work). The method is robust and converges rapidly.[33] Choosing initial conditions to be the solution for an S-system with terms selected in a combinatorial manner from among the terms of the larger GMA-system has been shown to find many, and in some cases all, of the roots for the original equations.[34]

3. Mathematically controlled comparison of alternatives

The existence of an explicit solution allows for the analytical specification of systemic constraints or invariants that provide the basis for the method of mathematically controlled comparisons.[10,11,27,30,35,36] The method involves the following steps. (1) The alternatives being compared are restricted to having differences in a single specific process that remains embedded within its natural milieu. (2) The values of the parameters that characterize the unaltered processes of one system are assumed to be strictly identical with those of the corresponding parameters of the alternative system. This equivalence of parameter values within the systems is called *internal equivalence*. It provides a means of nullifying or diminishing the influence of the background, which in complex systems is largely unknown. (3) Parameters associated with the changed process are initially free to assume any value. This allows the creation of new degrees of freedom. (4) The extra degrees of freedom are then systematically reduced by imposing constraints on the external behavior of the systems, e.g., by insisting that signals transmitted from input (independent variables) to output (dependent variables) be amplified by the same factor in the alternative systems. In this way the two systems are made as nearly equivalent as possible in their interactions with the outside environment. This is called *external equivalence*. (5) The constraints imposed by external equivalence fix the values of the altered parameters in such a way that arbitrary differences in systemic behavior are eliminated. Functional differences that remain between alternative systems with maximum internal and external equivalence constitute irreducible differences. (6) When all degrees of freedom have been eliminated, and the alternatives are as close to equivalent as they can be, then comparisons are made by rigorous mathematical and computer analyses of the alternatives.

Two key features of this method should be noted. First, because much of the analysis can be carried out symbolically, the results are often independent of the numerical values for particular parameters. This is a marked advantage because one does not know, and in many cases it would be impractical to obtain, all the parameter values of a complex system. Second, the method allows one to determine the relative optima of alternative designs without actually having to carry out an optimization (i.e., without having to determine explicit values for the parameters that optimize the performance of a given design). If one can show that a given design with an arbitrary set of parameter values is always superior to the alternative design that has been made internally and externally equivalent, whether or not the set of parameter values represents an optimum for either design, then one has proved that the given design will be superior to the alternative even if the alternative were assigned a parameter set that optimized its performance. This feature is a decided advantage because one can avoid the difficult procedure of optimizing complex nonlinear systems.

The method of mathematically controlled comparison has been used for some time to determine which of two alternative regulatory designs is better according to specific quantitative criteria for functional effectiveness. In some

Downloaded 20 Mar 2001 to 141.214.51.241. Redistribution subject to AIP copyright, see http://ojps.aip.org/chaos/chocr.jsp

cases, as noted above, the results obtained using this technique are general and qualitatively clear cut, i.e., one design is always better than another, independent of parameter values. For example, consider some systemic property, say a particular parameter sensitivity, whose magnitude should be as small as possible. In many cases, the ratio of this property in the alternative design relative to that in the reference design has the form $R=A/(A+B)$ where A and B are positive quantities with a distinct composition involving many individual parameters. Such a ratio is always less than one, which indicates that the alternative design is superior to the reference design with regard to this desirable property. In other cases, the results might be general but difficult to demonstrate because the ratio has a more complex form, and comparisons made with specific values for the parameters can help to clarify the situation. In either of these cases, comparisons made with specific values for the parameters also can provide a quantitative answer to the question of how much better one of the alternatives might be.

In contrast to the cases discussed previously, in which a clear-cut qualitative difference exists, a more ambiguous result is obtained when either of the alternatives can be better, depending on the specific values of the parameters. For example, the ratio of some desirable systemic property in the alternative design relative to that in the reference design has the form $R=(A+C)/(A+B)$, where A, B, and C are positive quantities with a distinct composition involving many individual parameters. For some values of the individual parameters $C>B$ and for other values $C<B$, so there is no clear-cut qualitative result. A numerical approach to this problem has recently been developed that combines the method of mathematically controlled comparison with statistical techniques to yield numerical results that are general in a statistical sense.[37] This approach retains some of the generality that makes mathematically controlled comparison so attractive, and at the same time provides quantitative results that are lacking in the qualitative approach.

IV. EXAMPLES OF DESIGN PRINCIPLES FOR ELEMENTARY GENE CIRCUITS

Each design feature of gene circuits allows for several differences in design. Our goal is to discover the design principles, if such exist, that would allow one to make predictions concerning which of the different designs would be selected under various conditions. For most features, the design principles are unknown, and we are currently unable to predict which design among a variety of well-characterized designs might be selected in a given context. In a few cases, as reviewed later, principles have been uncovered. There are simple rules that predict whether the mode of control will be positive or negative, whether elementary circuits will be directly coupled, inversely coupled, or uncoupled, and whether gene expression will switch in a static or dynamic fashion. More subtle conditions relate the logic of gene expression to the context provided by the life cycle of the organism.

A. Molecular mode of control

A simple demand theory based on selection allows one to predict the molecular mode of gene control. This theory states that the mode of control is correlated with the demand for gene expression in the organism's natural environment: positive when demand is high and negative when demand is low. Development of this theory involved elucidating functional differences, determining the consequences of mutational entropy (the tendency for random mutations to degrade highly ordered structures rather than contribute to their formation), and examining selection in alternative environments.

TABLE I. Predicted correlation between molecular mode of control and the demand for gene expression in the natural environment.

	Mode of control	
Demand for expression	Positive	Negative
High	Selected	Lost
Low	Lost	Selected

Detailed analysis involving mathematically controlled comparisons demonstrates that model gene circuits with the alternative modes of control behave identically in most respects. However, they respond in diametrically opposed ways to mutations in the control elements themselves.[27] Mutational entropy leads to loss of control in each case. However, this is manifested as super-repressed expression in circuits with the positive mode of control, and constitutive expression in circuits with the negative mode. The dynamics of mixed populations of organisms that harbor either the mutant or the wild-type control mechanism depend on whether the demand for gene expression in the environment is high or low.[38] The results are summarized in Table I. The basis for these results can be understood in terms of the following qualitative argument involving extreme environments.

A gene with a positive mode of control and a high demand for its expression will be induced normally if the control mechanism is wild type. It will be uninduced if the control mechanism is mutant, and, since expression cannot meet the demand in this case, the organism harboring the mutant mechanism will be selected against. In other words, the functional positive mode of control will be selected when mutant and wild-type organisms grow in a mixed population. On the other hand, in an environment with a low demand for expression, the gene will be uninduced in both wild-type and mutant organisms and there will be no selection. Instead, the mutants will accumulate with time because of mutational entropy, and the wild-type organisms with the functional positive mode of control will be lost.

The results for the negative mode of control are just the reverse. A gene with a negative mode of control and a low demand for its expression will be uninduced normally if the control mechanism is wild type. It will be constitutively induced if the control mechanism is mutant, and, since inappropriate expression in time or space tends to be dysfunctional, the organism harboring the mutant mechanism will be selected against. In other words, the functional negative mode of control will be selected when mutant and wild-type organisms grow in a mixed population. On the other hand, in an environment with a high demand for expression, the gene will be induced in both wild-type and mutant organisms and

Downloaded 20 Mar 2001 to 141.214.51.241. Redistribution subject to AIP copyright, see http://ojps.aip.org/chaos/chocr.jsp

TABLE II. General predictions regarding the mode of control for regulation of cell-specific functions in differentiated cell types.[a]

Cell type	Cell-specific functions	
	A	B
A	Positive	Negative
B	Negative	Positive

[a]See Fig. 7 and discussion in the text for a specific example.

there will be no selection. Instead, the mutants will accumulate with time because of mutational entropy, and the wild-type organisms with the functional negative mode of control will be lost.

The predictions of demand theory are in agreement with nearly all individual examples for which both the mode of control and the demand for expression are well-documented.[39] On the basis of this strong correlation, one can make predictions concerning the mode of control when the natural demand for expression is known, or vice versa. Moreover, when knowledge of cellular physiology dictates that pairs of regulated genes should be subject to the same demand regime, even if it is unknown whether the demand in the natural environment is high or low, then demand theory allows one to predict that the mode of control will be of the same type for both genes. Conversely, when such genes should be subject to opposite demand regimes, and again even if it is unknown whether the demand in the natural environment is high or low, then demand theory allows one to predict that the mode of control will be of the opposite type for these genes. The value of such predictions is that once the mode of control is determined experimentally for one of the two genes, one can immediately predict the mode of control for the other.

Straightforward application of demand theory to the control of cell-specific functions in differentiated cell types not only makes predictions about the mode of control for these functions in each of the cell types, but also makes the surprising prediction that the mode of control itself ought to undergo switching during differentiation from one cell type to another.[40] Table II summarizes the general predictions, and Fig. 7 provides a specific example of a simple model system, cells of *Escherichia coli* infected with the temperate bacteriophage λ, that fulfills these predictions. During lytic growth (cell type A in Table II), the lytic functions (A-specific functions) are in high demand and are predicted to involve the positive mode of control. Indeed, they are controlled by the N gene product, which is an anti-terminator exercising a positive mode of control. At the same time, the lysogenic functions (B-specific functions) are in low demand and are predicted to involve the negative mode of control. In this case, they are controlled by the CRO gene product, which is a repressor exercising a negative mode of control. Conversely, during lysogenic growth (cell type B in Table II), the lytic functions (A-specific functions) are in low demand and are predicted to involve the negative mode of control. Indeed, they are controlled by the CI gene product, which is a repressor exercising a negative mode of control. At the same time, the lysogenic functions (B-specific functions) are in high demand and are predicted to involve the positive mode of control. In this case, they are controlled by the CI gene product, which is also an activator exercising a positive mode of control. The mode in each individual case is predicted correctly, and the switching of modes during "differentiation" (from lysogenic to lytic growth or vice versa) is brought about by the interlocking circuitry of phage λ.

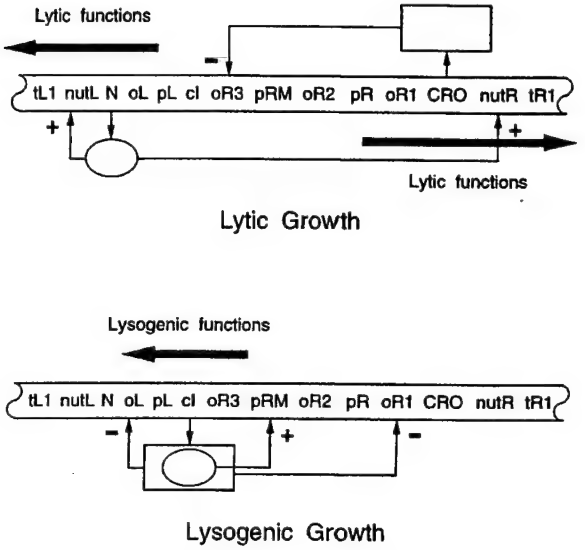


FIG. 7. Switching the mode of control for regulated cell-specific functions during differentiation. The temperate bacteriophage λ can be considered a simple model system that exhibits two differentiated forms: (Top panel) The lytic form in which the phage infects a cell, multiplies to produce multiple phage copies, lysis the cell, and the released progeny begin another cycle of lytic growth. (Bottom panel) The lysogenic form in which the phage genome is stably incorporated into the host cell DNA and is replicated passively once each time the host genome is duplicated. During differentiation, when the lysogenic phage is induced to become a lytic phage or the lytic phage becomes a lysogenic phage upon infection of a bacterial cell, the mode of control switches from positive to negative or vice versa because of the interlocking gene circuitry of phage λ. See text for further discussion.

B. Coupling of elementary gene circuits

There are logically just three patterns of coupling between the expression cascades for regulator and effector proteins in elementary gene circuits. These are the directly coupled, uncoupled, and inversely coupled patterns in which regulator gene expression increases, remains unchanged, or decreases with an increase in effector gene expression (Fig. 8). Elementary gene circuits in bacteria have long been studied and there are well-characterized examples that exhibit each of these patterns.

A design principle governing the pattern of coupling in such circuits has been identified by mathematically controlled comparison of various designs.[11] The principle is expressed in terms on two properties: the mode of control (positive or negative) and the capacity for regulated expression (large or small ratio of maximal to basal level of expression). According to this principle, one predicts that elementary gene circuits with the negative mode and small, intermediate, and large capacities for gene regulation will

Downloaded 20 Mar 2001 to 141.214.51.241. Redistribution subject to AIP copyright, see http://ojps.aip.org/chaos/chocr.jsp

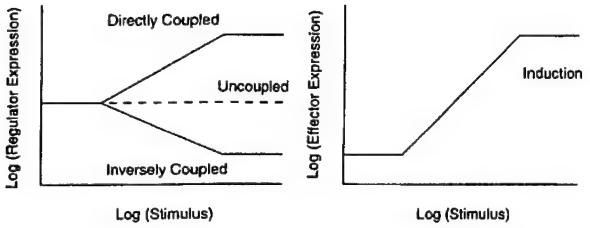


FIG. 8. Coupling of expression in elementary gene circuits. The panel on the right shows the steady-state expression characteristic for the effector cascade in Fig. 6. The panel on the left shows the steady-state expression characteristic for the regulator cascade. Induction of effector expression occurs while regulator expression increases (directly coupled expression), remains unchanged (uncoupled expression), or decreases (inversely coupled expression).

exhibit direct coupling, uncoupling, and inverse coupling, respectively. Circuits with the positive mode, in contrast, are predicted to have inverse coupling, uncoupling, and direct coupling.

The approach used to identify this design principle involves (1) formulating kinetic models that are sufficiently generic to include all of the logical possibilities for coupling of expression in elementary gene circuits, (2) making these models equivalent in all respects other than their regulatory parameters, (3) identifying a set of *a priori* criteria for functional effectiveness of such circuits, (4) analyzing the steady-state and dynamic behavior of the various designs, and (5) comparing the results to determine which designs are better according to the criteria. These steps are outlined next.

The kinetic models are all special cases of the generic model that is graphically depicted in Fig. 6. This model, which captures the essential features of many actual circuits, includes two transcription units: one for a regulator gene and another for a set of effector genes. The regulator gene encodes a protein that acts at the level of transcription to bring about induction, and the effector genes encode the enzymes that catalyzes a pathway of reactions in which the inducer is an intermediate. The regulator can negatively or positively influence transcription at the promoter of each transcription unit, and these influences, whether negative or positive, can be facilitated or antagonized by the inducer. A local power-law representation that describes the regulatable region (i.e., the inclined portion) of the steady-state expression characteristics in Fig. 8 is the following:

$$dX_1/dt = \alpha_1 X_3^{g13} X_5^{g15} X_6^{g16} - \beta_1 X_1^{h11}, \quad (4)$$

$$dX_2/dt = \alpha_2 X_1^{g21} X_7^{g27} - \beta_2 X_2^{h22}, \quad (5)$$

$$dX_3/dt = \alpha_3 X_2^{g32} X_8^{g38} - \beta_3 X_2^{h32} X_3^{h33}, \quad (6)$$

$$dX_4/dt = \alpha_4 X_3^{g43} X_5^{g45} X_6^{g46} - \beta_4 X_4^{h44}, \quad (7)$$

$$dX_5/dt = \alpha_5 X_4^{g54} X_7^{g57} - \beta_5 X_5^{h55}. \quad (8)$$

There are four parameters that characterize the pattern of regulatory interactions: g_{13} and g_{43} quantify influences of inducer X_3 on the rate of synthesis of effector mRNA X_1 and regulator mRNA X_4, whereas g_{15} and g_{45} quantify influences of regulator X_5 on these same processes.

TABLE III. Predicted patterns of coupling for regulator and effector cascades in elementary gene circuits.

Mode of control	Capacity for regulation[a]	Pattern of coupling
Positive	Large	Directly coupled
Positive	Intermediate	Uncoupled
Positive	Small	Inversely coupled
Negative	Large	Inversely coupled
Negative	Intermediate	Uncoupled
Negative	Small	Directly coupled

[a]Capacity for regulation is defined as the ratio of maximal to basal level of expression.

In the various models, the values for all corresponding parameters other than the four regulatory parameters are made equal (internal equivalence). The four regulatory parameters have their values constrained so as to produce the same steady-state expression characteristics (external equivalence). Models exhibiting each of the three patterns of coupling are represented within the space of the constrained regulatory parameters.

Six quantitative, *a priori* criteria for functional effectiveness are used as a basis for comparing the behavior of the various models. These are decisiveness, efficiency, selectivity, stability, robustness, and responsiveness. A decisive system has a sharp threshold for response to substrate. An efficient system makes a large amount of product from a given supra-threshold increment in substrate. A selective system governs the amount of regulator so as to ensure specific control of effector gene expression. A locally stable system returns to its original state following a small perturbation. A robust system tends to maintain its state despite changes in parameter values that determine its structure. A responsive system quickly adjusts to changes. (Further discussion of these criteria and the means by which they are quantified can be found elsewhere.[11])

The steady-state and dynamic behavior of the various models is analyzed by standard algebraic and numerical methods, and the results are quantified according to the above criteria. Temporal responsiveness is a distinguishing criterion for effectiveness of these circuits. A comparison of results for models with the various patterns of coupling leads to the predicted correlations summarized in Table III.

To test these predicted correlations we identified 32 elementary gene circuits for which the mode of control was known and for which quantitative data regarding the capacities for regulator and effector gene expression were available in the literature. A plot of these data in Fig. 9 shows reasonable agreement with the predicted positive slope for the points representing circuits with a positive mode and the predicted negative slope for the points representing circuits with a negative mode. Global experiments that utilize microarray technology could provide more numerous and potentially more accurate tests of these predictions.

C. Connectivity and switching

Gene expression can be switched ON (and OFF) in either a discontinuous dynamic fashion or a continuously variable static fashion in response to developmental or environ-

Downloaded 20 Mar 2001 to 141.214.51.241. Redistribution subject to AIP copyright, see http://ojps.aip.org/chaos/chocr.jsp

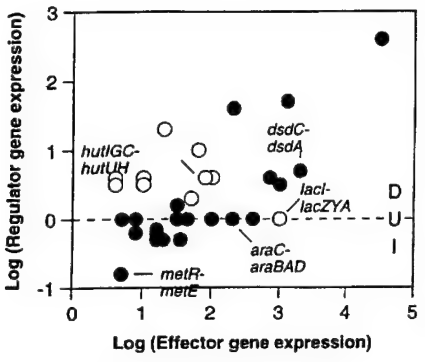


FIG. 9. Experimental data for the coupling of expression in elementary gene circuits. The capacity for induction of the effector cascade is plotted on the horizontal axis as positive values. The capacity for expression of the regulator cascade is plotted on the vertical axis as positive values (induction), negative values (repression), or zero (no change in expression). Effector cascades having a positive mode of control are represented as data points with filled symbols and those having a negative mode with open symbols. Data show reasonably good agreement with the predictions in Table III.

mental cues. These alternative switch behaviors are clearly manifested in the steady-state expression characteristic of the gene. In some cases, the elements of the circuitry appear to be the same, and yet the alternative behaviors can be generated by the way in which the elements are interconnected. This design feature has been examined in simple model circuits. The results have led to specific conditions that allow one to distinguish between these alternatives, and these conditions can be used to interpret the results of experiments with the *lac* operon of *E. coli*.

A design principle that distinguishes between discontinuous and continuous switches in a model for inducible catabolic pathways (Fig. 10) is the following. If the natural inducer is the initial substrate of the inducible pathway, or if it is an intermediate in the inducible pathway, then the switch will be continuous; if the inducer is the final product of the inducible pathway, then the switch can be discontinuous or continuous, depending on an algebraic condition that involves four kinetic orders for reactions in the circuit. (A more general statement of the principle can be given in terms of the algebraic condition, as will be shown below.)

A simplified set of equations that captures the essential features of the model in Fig. 10 is the following:

$$dX_1/dt=\alpha_{1B}-\beta_1 X_1, \quad X_3<X_{3L}, \tag{9a}$$

$$dX_1/dt=\alpha_1 X_3^{g_{13}}-\beta_1 X_1, \quad X_{3L}<X_3<X_{3H}, \tag{9b}$$

$$dX_1/dt=\alpha_{1M}-\beta_1 X_1, \quad X_{3H}<X_3, \tag{9c}$$

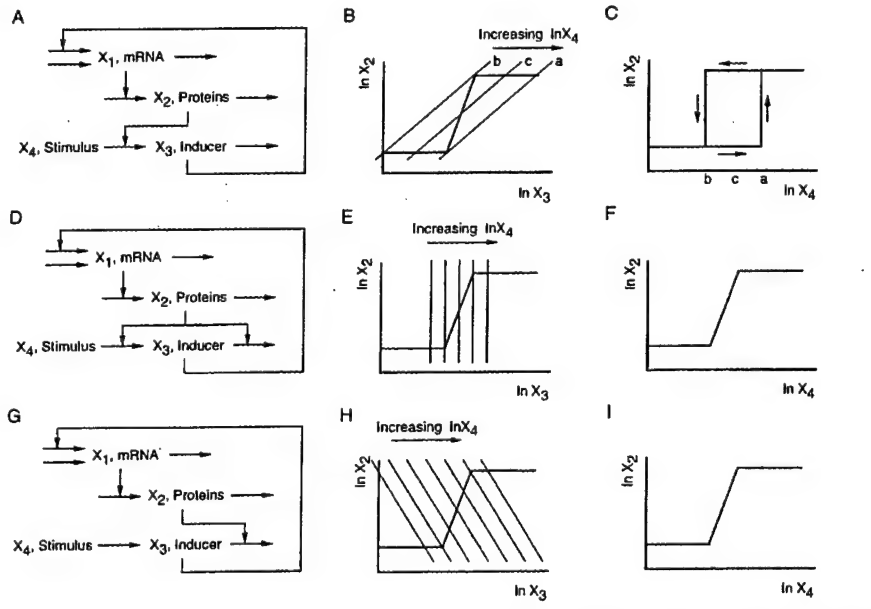


FIG. 10. Simplified model of an inducible catabolic pathway exhibits two types of switch behavior depending upon the position of the inducer in the inducible pathway. (a) The inducer is the final product of the inducible pathway. (b) The S-shaped curve is the steady-state solution for Eqs. (9) and (10). The lines (*a*, *b*, and *c*) are the steady-state solutions for Eq. (11) with different fixed concentrations of the stimulus X_4. The steady-state solutions for the system are given by the intersections of the S-shaped curve and the straight lines. There is only one intersection (maximal expression) when $\ln X_4>a$; there is only one intersection (basal expression) when $\ln X_4<b$. There are three intersections when $b<\ln X_4=c<a$, but the middle one is unstable. The necessary and sufficient condition for the bistable behavior in this context is that the slope of the straight line be less than the slope of the S-shaped curve at intermediate concentrations of the inducer X_3, which is the condition expressed in Eq. (12). (c) The steady-state induction characteristic exhibits discontinuous dynamic switches and a well-defined hysteresis loop. Thus, at intermediate concentrations of the stimulus X_4, expression will be at either the maximal or the basal level depending upon the past history of induction. (d) The inducer is an intermediate in the inducible pathway. (e) The steady-state solutions for the system are given by the intersections of the S-shaped curve and vertical lines. There is only one intersection possible for any given concentration of stimulus. (f) The steady-state induction characteristic exhibits a continuously changing static switch. (g) The inducer is the initial substrate of the inducible pathway. (h) The steady-state solutions for the system are given by the intersections of the S-shaped curve and the lines of negative slope. There is only one intersection possible for any given concentration of stimulus. (i) The steady-state induction characteristic exhibits a continuously changing static switch.

Downloaded 20 Mar 2001 to 141.214.51.241. Redistribution subject to AIP copyright, see http://ojps.aip.org/chaos/chocr.jsp

TABLE IV. Summary of predictions relating type of switch behavior to the connectivity in the model inducible circuit of Fig. 10.

Figure	Stimulus	Inducer	Transport	Connection from inducible pathway	Switch
10(d)	IPTG	IPTG	Constitutive	None	Static
10(a)	IPTG	IPTG	Inducible	Product	Dynamic
10(d)	Lactose	Allolactose	Inducible	Intermediate	Static
10(g)	Allolactose	Allolactose	Constitutive	Substrate	Static

$$dX_2/dt=\alpha_2 X_1-\beta_2 X_2, \tag{10}$$

$$dX_3/dt=\alpha_3 X_2^{g32} X_4^{g34}-\beta_3 X_2^{h32} X_3^{h33}. \tag{11}$$

The variables X_1, X_2, X_3, and X_4 represent the concentrations of polycistronic mRNA, a coordinately regulated set of proteins, inducer, and stimulus, respectively. This is a piecewise power-law representation (see Appendix of Ref. 27) that emphasizes distinct regions of operation. There is a constant basal level of expression when inducer concentration X_3 is lower than a value X_{3L}; there is a constant maximal level of expression when inducer concentration is higher than a value X_{3H}; there is a regulated level of expression (with cooperativity indicated by a value of the parameter $g_{13}>1$) when inducer concentration is between the values X_{3L} and X_{3H}. All parameters in this model have positive values.

The position of the natural inducer in an inducible pathway has long been known to have a profound effect on the local stability of the steady state when the system is operating on the inclined portion (i.e., the regulatable region) of the steady-state expression characteristic (Fig. 8, right panel).[27] As the position of the natural inducer is changed from the initial substrate [Fig. 10(g)], to an intermediate [Fig. 10(d)] to the final product [Fig. 10(a)] of the inducible pathway (all other parameters having fixed values), the margin of stability decreases. In this progression the single stable steady state [Fig. 10(h)] can undergo a bifurcation to an unstable steady state flanked by two stable steady states [Fig. 10(b)], which is the well-known cusp catastrophe characteristic of a dynamic ON–OFF switch.[41]

The critical conditions for the existence of multiple steady states and a dynamic switch are given by

$$g_{13}>h_{33}/(g_{32}-h_{32}) \quad \text{and} \quad g_{32}>h_{32}. \tag{12}$$

In general, the inducible proteins must have a greater influence on the synthesis (g_{32}) than on the degradation (h_{32}) of the inducer. These conditions can be interpreted, according to conventional assumptions, in terms of inducer position in the pathway. If the position of the true inducer is functionally equivalent to that of the substrate for the inducible pathway, then $g_{32}=0$ and the conditions in Eq. (12) cannot be satisfied. If the position is functionally equivalent to that of the intermediate in the inducible pathway, the kinetic orders for the rates of synthesis and degradation of the intermediate are the same with respect to the enzymes for synthesis and degradation, and these enzymes are coordinately induced, then $g_{32}=h_{32}$ and again the conditions in Eq. (12) cannot be satisfied. However, if the position is functionally equivalent to that of the product for the inducible pathway, then $h_{32}=0$ and the conditions in Eq. (12) can be satisfied provided $g_{13}>h_{33}/g_{32}$.

The values of the parameters in this model have been estimated from experimental data for the *lac* operon of *E. coli*.[10] These results, together with these data, can be used to interpret four experiments involving the circuitry of the *lac* operon (see Table IV and the following discussion).

First, if the *lac* operon is induced with the nonmetabolizable (gratuitous) inducer isopropyl-β, D-thiogalactoside (IPTG) in a cell with the inducible Lac permease protein, then the model is as shown in Fig. 10(a). In this case, X_1 is the concentration of polycistronic *lac* mRNA, X_2 is the concentration of the Lac permease protein alone (X_2 has no influence on the degradation of the inducer X_3), X_3 is the intracellular concentration of IPTG, and X_4 is the extracellular concentration of IPTG. With the parameter values from the *lac* operon, the conditions in Eq. (12) are satisfied because $h_{33}=1$ (aggregate loss by all causes in exponentially growing cells is first order), $g_{32}=1$ (enzymatic rate is first order with respect to the concentration of total enzyme), and $g_{13}=2$ (the Hill coefficient of *lac* transcription with respect to the concentration of inducer is second order).

Second, if the *lac* operon is induced with the gratuitous inducer IPTG in a cell without the Lac permease protein, then the inducer IPTG is not acted upon by any of the protein products of the operon. In this case, X_1 is the concentration of polycistronic *lac* mRNA, X_2 is the concentration of β-galactosidase protein alone (X_2 has no influence on either the synthesis or the degradation of the inducer X_3), X_3 is the intracellular concentration of IPTG, and X_4 is the extracellular concentration of IPTG. The conditions in Eq. (12) now cannot be satisfied because $g_{23}=h_{23}=0$ and all other parameters are positive. This is an open-loop situation in which expression of the operon is simply proportional to the rate of transcription as determined by the steady-state concentration of intracellular IPTG, which is proportional to the concentration of extracellular IPTG.

Thus, the kinetic model accounts for two important observations from previous experiments. It accounts for the classic experimental results of Novick and Weiner[42] in which they observed a discontinuous dynamic switch with hysteresis. They induced the *lac* operon with a gratuitous inducer that was transported into the cell by the inducible Lac permease, was diluted by cellular growth, but was not acted upon by the remainder of the inducible pathway. Hence, the gratuitous inducer occupied the position of final product for the inducible pathway (in this case simply the Lac permease

Downloaded 20 Mar 2001 to 141.214.51.241. Redistribution subject to AIP copyright, see http://ojps.aip.org/chaos/chocr.jsp

step), and the model predicts dynamic bistable switch behavior similar to that depicted in Figs. 10(b) and 10(c). The kinetic model also accounts for the classic experimental results of Sadler and Novick[43] in which they observed a continuous static switch without hysteresis. In their experiments they used a mutant strain of *E. coli* in which the *lac* permease was inactivated and they induced the *lac* operon with a gratuitous inducer. In this system, the inducer is not acted upon by any part of the inducible pathway, the extracellular and intracellular concentrations of inducer are proportional, and the model predicts a continuous static switch similar to that depicted in Figs. 10(e) and 10(f). The model in Fig. 10 also makes two other predictions related to the position of the natural inducer in the inducible pathway.

First, if the *lac* operon is induced with lactose in a cell with all the inducible Lac proteins intact, then the model is as shown in Fig. 10(d). In this case, X_1 is the concentration of polycistronic *lac* mRNA, X_2 is the concentration of the Lac permease protein as well as the concentration of the β-galactosidase protein (which catalyzes both the conversion of lactose to allolactose and the conversion of allolactose to galactose and glucose), X_3 is the intracellular concentration of allolactose, and X_4 is the extracellular concentration of lactose. In steady state, the sequential conversion of extracellular lactose to intracellular lactose (by Lac permease) and intracellular lactose to allolactose (by β-galactosidase) can be represented without loss of generality as a single process because these two proteins are coordinately expressed. Again, the conditions in Eq. (12) cannot be satisfied. In this case, $g_{23}=h_{23}=1$ and all other parameters are positive finite, and the model predicts a continuous static switch similar to that depicted in Figs. 10(e) and 10(f).

Second, if the *lac* operon is induced with allolactose, the natural inducer, in a cell without the Lac permease protein, then the model is as shown in Fig. 10(g). In this case, X_1 is the concentration of polycistronic *lac* mRNA, X_2 is the concentration of the β-galactosidase protein alone (which catalyzes the conversion of allolactose to galactose and glucose), X_3 is the intracellular concentration of allolactose, and X_4 is the extracellular concentration of allolactose. The conditions in Eq. (12) cannot be satisfied. In this case, $g_{23}=0$ and all other parameters are positive, and the model predicts a continuous static switch similar to that represented in Figs. 10(h) and 10(i).

The fact that the kinetic model of the *lac* operon predicts a continuous static switch in response to extracellular lactose led us to search the literature for the relevant experimental data. We were unable to find any experimental evidence for either a continuous static switch or a discontinuous dynamic switch in response to lactose, which comes as a surprise. Despite the long history of study involving the *lac* operon, such experiments apparently have not been reported. Experiments to test this prediction specifically are currently being designed and carried out (Atkinson and Ninfa, unpublished results).

D. Context and logic

In the qualitative version of demand theory (Sec. IV A) it was assumed for simplicity that there was a constant demand regime for the effector gene in question and that its expression was controlled by a single regulator. Here I review the quantitative version of demand theory and include consideration of genes exposed to more than one demand regime and controlled by more than one regulator.

1. Life cycle provides the context for gene control

Models that include consideration of the organism's life cycle, molecular mode of gene control, and population dynamics are used to describe mutant and wild-type populations in two environments with different demands for expression of the genes in question. These models are analyzed mathematically in order to identify conditions that lead to either selection or loss of a given mode of control. It will be shown that this theory ties together a number of important variables, including growth rates, mutation rates, minimum and maximum demands for gene expression, and minimum and maximum durations for the life cycle of the organism. A test of the theory is provided by the *lac* operon of *E. coli*.

The life cycle of *E. coli* involves sequential colonization of new host organisms,[44] which means repeated cycling between two different environments [Figs. 11(a) and 11(b)]. In one, the upper portion of the host's intestinal track, the microbe is exposed to the substrate lactose and there is a high demand for expression of the *lac* operon, and in the other, the lower portion of the intestinal track and the environment external to the host, the microbe is not exposed to lactose and there is a low demand for *lac* expression. The average time to complete a cycle through these two environments is defined as the cycle time, C, and the average fraction of the cycle time spent in the high-demand environment is defined as the demand for gene expression, D [Fig. 11(c)].

The implications for gene expression of mutant and wild-type operons in the high- and low-demand environments are as follows. The wild-type functions by turning on expression in the high-demand environment and turning off expression in the low-demand environment. The mutant with a defective promoter is unable to turn on expression in either environment. The mutant with a defective modulator (or defective regulator protein) is unable to turn off expression regardless of the environment. The double mutant with defects in both promoter and modulator/regulator behaves like the promoter mutant and is unable to turn on expression in either environment.

The sizes of the populations are affected by the transfer rate between populations, which is the result of mutation, and by the growth rate, which is the result of overall fitness. The transfer rates depend on the mutation rate per base and on the size of the relevant target sequence. The growth rate for the wild type is greater than that for mutants of the modulator type in the low-demand environment; these mutants are selected against because of their superfluous expression of an unneeded function. The growth rate for the wild type is greater than that for mutants of the promoter type in the high-demand environment; these mutants are selected against because of their inability to express the needed function.

Solution of the dynamic equations for the populations cycling through the two environments yields expressions in

Downloaded 20 Mar 2001 to 141.214.51.241. Redistribution subject to AIP copyright, see http://ojps.aip.org/chaos/chocr.jsp

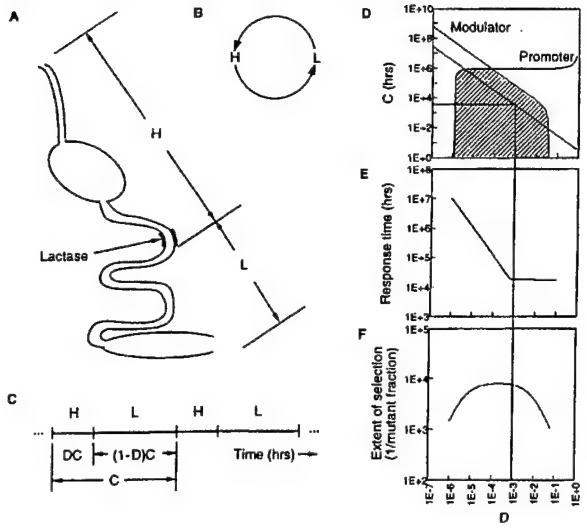


FIG. 11. Life cycle of *Escherichia coli* and the demand for expression of its *lac* operon. (a) Schematic diagram of the upper (high demand) and lower (low demand) portions of the human intestinal track. (b) Life cycle consists of repeated passage between environments with high- and low-demand for *lac* gene expression. (c) Definition of cycle time C and demand for gene expression D. (d) Region in the C vs D plot for which selection of the wild-type control mechanism is possible. (e) Rate of selection as a function of demand. (f) Extent of selection as a function of demand. See text for discussion.

C and D for the threshold, extent, and rate of selection that apply to the wild-type control mechanism.[45] The threshold for selection is given by the boundary of the shaded region in Fig. 11(d); only systems with values of C and D that fall within this region are capable of being selected. The rate and extent of selection shown in Figs. 11(e) and 11(f) exhibit optimum values for a specific value of D.

Application of this quantitative demand theory to the *lac* operon of *E. coli* yields several new and provocative predictions that relate genotype to phenotype.[46] The straight line in Fig. 11(d) represents the inverse relationship $C=3/D$ that results from fixing the time of exposure to lactose at 3 h, which is the clinically determined value for humans.[47,48] The intersections of this line with the two thresholds for selection provide lower and upper bounds on the cycle time. The lower bound is approximately 24 h, which is about as fast as the microbe can cycle through the intestinal track without colonization.[49–51] The upper bound is approximately 70 years, which is the longest period of colonization without cycling and corresponds favorably with the maximum life span of the host.[52] The optimum value for the cycle time, as determined by the optimum value for demand [from Figs. 11(e) and 11(f)], is approximately four months, which is comparable to the average rate of recolonization measured in humans.[53–55] A summary of these results is given in Table V.

2. Logic unit and phasing of lac control

The analysis in Sec. IV D 1 assumed that when *E. coli* was growing on lactose there was no other more preferred carbon source present. Thus, the positive CAP-cAMP regulator[56] was always present, and we could then concentrate on the conditions for selection of the specific control by Lac repressor. This was a simplifying assumption; in the more general situation, both the specific control by Lac repressor and the global control by CAP-cAMP activator must be taken into consideration. The analysis becomes more complex, but it follows closely the outline of the simpler case in Sec. IV D 1.

TABLE V. Summary of experimental data and model predictions based on conditions for selection of the *lac* operon in *Escherichia coli*.

Characteristic	Experimental data	Model predictions
Intestinal transit time	12–48 h	26 h
Lifetime of the host	120 years	66 years
Re-colonization rate	2–18 months	4 months

By extension of the definition for demand D, given in Sec. IV D 1, one can define a period of demand for the absence of repressor G, a period of demand for the presence of activator E, and a phase relationship between these two periods of demand F. By extension of the analysis in Sec. IV D 1, solution of the dynamic equations for wild-type and mutant populations cycling through the two environments yields expressions in C, G, E, and F for the threshold, extent, and rate of selection that apply to the wild-type control mechanism.

The threshold for selection is now an envelope surrounding a "mound" in four-dimensional space with cycle time C as a function of the three parameters G, E, and F; only systems with values that fall within this envelope are capable of being selected. The rate and extent of selection exhibit optimum values as before, but these now occur with a specific combination of values for G, E, and F. The values of G, E, and F that yield the optima represent a small period when repressor is absent, an even smaller period when activator is absent, and a large phase period between them. The period when repressor is absent corresponds to the period of exposure to lactose (~0.36% of the cycle time). Within this period (but shifted by ~0.20% of the cycle time) there is a shorter period when activator is absent (~0.14% of the cycle time); this corresponds to the presence of a more preferred carbon source that lowers the level of cAMP.

These relationships can be interpreted in terms of exposure to lactose, exposure to glucose, and expression of the *lac* operon as shown in Fig. 12. As *E. coli* enters a new host, passes through the early part of the intestinal track, and is exposed to lactose, the *lac* operon is induced and the bacteria are able to utilize lactose as a carbon source. During this period the operator site of the *lac* operon is free of the Lac repressor. At the point in the small intestine where the host's lactase enzymes are localized, lactose is actively split into its constituent sugars, glucose and galactose. This creates a rapid elevation in the concentration of these sugars in the environment of *E. coli*. A period of growth on glucose is initiated, and this is accompanied by catabolite repression and lactose exclusion from the bacteria. During this period the initiator site of the *lac* operon is free of the CAP-cAMP activator, transcription of the *lac* operon ceases, and the concentration of β-galactosidase is diluted by growth. During

Downloaded 20 Mar 2001 to 141.214.51.241. Redistribution subject to AIP copyright, see http://ojps.aip.org/chaos/chocr.jsp

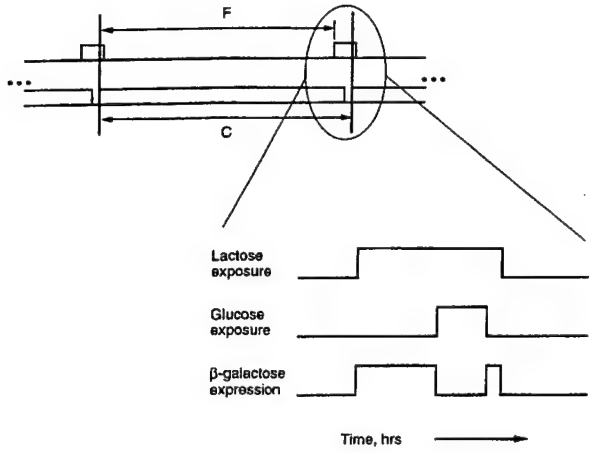


FIG. 12. Optimal duration and phasing of the action by the positive (CAP-cAMP) and negative (LacI) regulators of β-galactosidase expression. The signal on the top line represents the absence of repressor binding to the *lac* operator site, the signal on the second line represents activator binding to the *lac* initiator site. The cycle time C is the period between the vertical lines, and the relative phasing is shown as F. An expanded view of the critical region gives an interpretation in terms of exposure to lactose and glucose as bacteria pass the site of the lactase enzymes in the small intestine. See text for discussion.

this period the glucose in the intestine is also rapidly absorbed by the host. Eventually, the glucose is exhausted, the CAP-cAMP activator again binds the initiator site of the *lac* operon, and the residual lactose that escaped hydrolysis by the host's lactase enzymes causes a diminished secondary induction of the bacterial *lac* operon. Finally, the lactose is exhausted, the Lac repressor again binds the operator site of the *lac* operon, and the microbe enters the low-demand environment and colonizes the host.

The quantitative version of demand theory integrates information at the level of DNA (mutation rate, effective target sizes for mutation of regulatory proteins, promoter sites, and modulator sites), physiology (selection coefficients for superfluous expression of an unneeded function and for lack of expression of an essential function), and ecology (environmental context and life cycle) and makes rather surprising predictions connected to the intestinal physiology, life span, and recolonization rate of the host. There is independent experimental data to support each of these predictions.

Finally, when the logic of combined control by CAP-cAMP activator and Lac repressor was analyzed, we found an optimum set of values not only for the exposure to lactose, but also for the exposure to glucose and for the relative phasing between these periods of exposure. The phasing predicted is consistent with the spatial and temporal environment created by the patterns of disaccharide hydrolysis and monosaccharide absorption along the intestinal tract of the host.

V. DISCUSSION

Although biological principles that govern some variations in design have been identified (e.g., positive vs negative modes of control), there are other well-documented (and many not so well-documented) variations in design that still are not understood. For example, why is the positive mode of control in some cases realized with an activator protein that facilitates transcription of genes downstream of a promoter, and in other cases with an antiterminator protein that facilitates transcription of genes downstream of a terminator? There are many examples of each, but no convincing explanation for the difference. Thus, the elements of design and the variations I have reviewed in Sec. II provide only a start; there is much to be done in this area.

For the comparative analysis of alternative designs we require a formalism capable of representing diverse designs, tractable methods of analysis for characterizing designs, and a strategy for making well-controlled comparisons that reveal essential differences while minimizing extraneous differences. As reviewed in Sec. III, there are several arguments that favor the power-law formalism for representing a wide spectrum of nonlinear systems. In particular, the local S-system representation within this formalism not only provides reasonably accurate descriptions but also possesses a tractable structure, which allows explicit solutions for the steady state and efficient numerical solutions for the dynamics. Explicit steady-state solutions are used to make mathematically controlled comparisons. Constraining these solutions provides invariants that eliminate extra degrees of freedom, which otherwise would introduce extraneous differences into the comparison of alternatives. The ability to provide such invariants is one of the principle advantages of using the local S-system representation. Two other formalisms with this property are the linear representation and the Volterra–Lotka representation, which is equivalent to the linear representation for the steady state. However, these representations yield linear relations between variables in steady state, which is less appropriate for biological systems in which these relationships are typically nonlinear.

The utility of these methods for studying alternative designs ultimately will be determined by the degree to which their predictions are supported by experimental evidence. For this reason it is important that the methods consider an entire class of systems without specifying numerical values for the parameters, which often are unknown in any case. Predictions achieved with this approach then can be tested against numerous examples provided by members of the class. If the methods were to focus upon a single system with specific values for its parameters, then there would be only the one example to test any hypothesis that might be conceived. The symbolic approach also allows one to compare efficiently many alternatives including ones that no longer exist (and so values of their parameters will never be known), which often is the case in trying to account for the evolution of a given design, or that hypothetically might be brought into existence through genetic engineering. The four design principles reviewed in Sec. IV illustrate the types of results that have been obtained when the methods in Sec. III are applied to some of the elements of design described in Sec. II.

First, we examined the two modes of control in elementary gene circuits (Sec. IV A). Qualitative arguments and examples were used to demonstrate the validity of demand

Downloaded 20 Mar 2001 to 141.214.51.241. Redistribution subject to AIP copyright, see http://ojps.aip.org/chaos/chocr.jsp

theory for the regulator–modulator component of control mechanisms. The same approach also can be used to account for the alternative forms of the promoter component. In either case, the qualitative arguments are based on extreme cases where the demand is clearly high or low. One would like to quantify what is meant by demand, to know how high it must be to select for the positive mode of control or a low-level promoter, and to know how low it must be to select for the negative mode of control or a high-level promoter. The quantitative version of demand theory reviewed in Sec. IV D specifically addresses these issues.

Second, we examined the three patterns of coupling in elementary gene circuits (Sec. IV B). It was their dynamic properties that proved to be distinctive. Establishing the dynamic differences required efficient numerical solutions of the differential equations and a means to reduce the dimension of the search in order to explore fully the parameter space. The results in Sec. IV B illuminate an area of experimental work that needs greater attention. For example, the data in Fig. 9 were obtained from individual gene circuits as a result of labor-intensive studies designed for purposes other than quantitative characterization of the steady-state induction characteristics for effector and regulator cascades. The data often are sketchy and subject to large errors, particularly in the case of regulator proteins, which generally are expressed at very low levels. Genomic and proteomic approaches to the measurement of expression should provide data for a much larger number of elementary gene circuits. However, these approaches also have difficulty measuring low levels of expression, and so technical improvements will be needed before they will be able to quantify expression of regulator genes.

Third, we examined various forms of connectivity that link the inducer to the transcription unit for an inducible catabolic pathway and showed that two different types of switching behavior result (Sec. IV C). The analysis of *lac* circuitry in this regard focused attention on a long-standing misconception in the literature, namely, that *lac* operon expression normally is an all-or-none phenomenon. While continuously variable induction of the lactose operon might be appropriate for a catabolic pathway whose expression can provide benefits to the cell even when partially induced, a discontinuous induction with hysteresis might be more appropriate for major differentiation events that require a definite commitment at some point. The wider the hysteretic loop the greater the degree of commitment. The width of the loop tends to increase with a large capacity for induction (ratio of maximum to basal level of expression), high logarithmic gain in the regulatable region (high degree of cooperativity), and substrates for the enzymes in the pathway operating as near saturation as compatible with switching.

Fourth, we examined the context of gene expression and developed a quantitative version of demand theory (Sec. IV D). In addition to providing a quantitative measure of demand, the results define what high and low mean in terms of the level of demand required to select for the positive or the negative mode of control and for low- or high-level promoters. This analysis also predicted new and unexpected kinds of information, such as intestinal transit time, host lifetime, and recolonization rate. When the logic unit involving the two relevant regulators was included in the analysis it also yielded predictions for the relative phasing of the environmental cues involved in *lac* operon induction.

Is there anything common to these successful explanations of design that might be useful as a guide in exploring other variations in design? Two such features come to mind. First, each of the examples involved a limited number of possible variations on a theme: two modes of control, three patterns of coupling, two types of switches. This meant that only a small number of cases had to be analyzed and compared, which is a manageable task. If there had been many variations in each case, then one would have no hope of finding a simple underlying rule that could account for all the variations, and one might never have considered analyzing and comparing all of the possibilities. Second, each case could be represented by a set of simple equations whose structure allowed symbolic analysis (and exhaustive numerical analysis when necessary). This permitted the use of controlled mathematical comparisons, which led to the identification of clear qualitative differences in the behavior of the alternatives. Thus, it might prove fruitful in the future to look for instances where these features present themselves.

In this context, we must acknowledge the fundamental role of accident in generating the diversity that is the substrate for natural selection. Thus, there undoubtedly will be examples of recently generated variations in design for which there will be no rational explanation. Only in time will natural selection tend to produce designs that are shaped for specific functions and hence understandable in principle.

Finally, will the understanding of large gene networks require additional tools beyond those needed for elementary gene circuits? Although we have no general answer to this question, there are three points having to do with network connectivity, catalytic versus stoichiometric linkages, and time-scale separation that are worthy of comment.

First, the evidence suggests, at least for bacteria, that there are relatively few connections between elementary gene circuits (see Sec. II D). This probably explains the experimental success that has been obtained by studying the regulation of isolated gene systems. Had there been rich interactions among these gene systems, such studies might have been less fruitful. Low connectivity also suggests that the understanding of elementary circuits may largely carry over to their role in larger networks and that the same tools might be used to study larger networks.

Second, catalytic linkages between circuits are less problematic then stoichiometric linkages, at least for the analysis of steady-state behavior. Elementary circuits can be linked catalytically without their individual properties changing appreciably, because the molecules in one circuit acting catalytically on another circuit are not consumed in the process of interaction. Such a circuit can have a unilateral effect on a second circuit, without having its own behavior affected in the process. This permits a modular block-diagram treatment, which makes use of the results obtained for the individual circuits in isolation, to characterize the larger network. (This is analogous to the well-known strategy used by electronic engineers, who design operational amplifiers with

Downloaded 20 Mar 2001 to 141.214.51.241. Redistribution subject to AIP copyright, see http://ojps.aip.org/chaos/chocr.jsp

high impedance to insulate the properties of the modules being coupled.) On the other hand, elementary circuits that are linked stoichiometrically may not be treatable in this fashion, because the molecules in one circuit are consumed in the process of interacting with a second circuit. This is a much more intimate linkage that may require the two circuits to be studied as a whole. In either case, the dynamic properties are not easily combined in general because the circuits are nonlinear.

Third, the separation of time scales allows some elementary circuits to be represented by transfer functions consisting of a simple power-law function. (Allometric relationships are an example of this.) This is related to the telescopic property of the S-system representation that was mentioned in Sec. III B 1. This property allows a simple block-diagram treatment of the elementary circuits that operate on a fast time scale.

ACKNOWLEDGMENTS

This work was supported in part by U.S. Public Health Service Grant No. RO1-GM30054 from the National Institutes of Health and by U.S. Department of Defense Grant No. N00014-97-1-0364 from the Office of Naval Research.

[1] F. Jacob and J. Monod, "Genetic regulatory mechanisms in the synthesis of proteins," J. Mol. Biol. **3**, 318–356 (1961).
[2] J. E. Hoch and T. J. Silhavy, *Two-Component Signal Transduction* (American Society for Microbiology, Washington, DC, 1995).
[3] M. A. Savageau, "Are there rules governing patterns of gene regulation?" in *Theoretical Biology—Epigenetic and Evolutionary Order*, edited by B. C. Goodwin and P. T. Saunders (Edinburgh University Press, Edinburgh, 1989), pp. 42–66.
[4] S. A. Kauffman, *The Origins of Order: Self-Organization and Selection in Evolution* (Oxford University Press, New York, 1993).
[5] K. E. Kürten and H. Beer, "Inhomogeneous Kauffman models at the borderline between order and chaos," J. Stat. Phys. **87**, 929–935 (1997).
[6] F. R. Blattner, G. Plunkett III, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau, and Y. Shao, "The complete genome sequence of *Escherichia coli* K-12," Science **277**, 1453–1462 (1997).
[7] D. Thieffry, H. Salgado, A. M. Huerta, and J. Collado-Vides, "Prediction of transcriptional regulatory sites in the complete genome sequence of *Escherichia coli* K-12," Bioinformatics **14**, 391–400 (1998).
[8] J. R. Beckwith and E. Zipser, Eds., *The Lactose Operon* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1970).
[9] C. H. Yuh, H. Bolouri, and E. H. Davidson, "Genomic cis-regulatory logic: Experimental and computational analysis of a sea urchin gene," Science **279**, 1896–1902 (1998).
[10] W. S. Hlavacek and M. A. Savageau, "Subunit structure of regulator proteins influences the design of gene circuitry: Analysis of perfectly coupled and completely uncoupled circuits," J. Mol. Biol. **248**, 739–755 (1995).
[11] W. S. Hlavacek and M. A. Savageau, "Rules for coupled expression of regulator and effector genes in inducible circuits," J. Mol. Biol. **255**, 121–139 (1996).
[12] F. C. Neidhardt and M. A. Savageau, "Regulation beyond the operon," in *Escherichia coli and Salmonella: Cellular and Molecular Biology*, edited by F. C. Neidhardt *et al.* (ASM, Washington, DC, 1996), pp. 1310–1324.
[13] A. Arkin, J. Ross, and H. H. McAdams, "Stochastic kinetic analysis of developmental pathway bifurcation in phage λ-infected *Escherichia coli* cells," Genetics **149**, 1633–1648 (1998).
[14] M. A. Savageau, "Biochemical systems analysis II. The steady state solutions for an n-pool system using a power-law approximation," J. Theor. Biol. **25**, 370–379 (1969).
[15] M. A. Savageau, "Power-law formalism: A canonical nonlinear approach to modeling and analysis," in *World Congress of Nonlinear Analysts 92*, edited by V. Lakshmikantham (de Gruyter, Berlin, 1996), Vol. 4, pp. 3323–3334.
[16] M. A. Savageau and E. O. Voit, "Recasting nonlinear differential equations as S-systems: A canonical nonlinear form," Math. Biosci. **87**, 83–115 (1987).
[17] A. Salvador, "Synergism analysis of biochemical systems. I. Conceptual framework," Math. Biosci. **163**, 105–129 (2000).
[18] A. Salvador, "Synergism analysis of biochemical systems. II. Deviations from multiplicativity," Math. Biosci. **163**, 131–158 (2000).
[19] E. O. Voit and M. A. Savageau, "Accuracy of alternative representations for integrated biochemical systems," Biochemistry **26**, 6869–6880 (1987).
[20] A. Sorribas and M. A. Savageau, "Strategies for representing metabolic pathways within biochemical systems theory: Reversible pathways," Math. Biosci. **94**, 239–269 (1989).
[21] Y. Maki, D. Tominaga, M. Okamoto, W. Watanabe, and Y. Eguchi, "Development of a system for the inference of large scale genetic networks," Pacific Symposium on Biocomputing **6**, 446–458 (2000).
[22] M. A. Savageau, "Growth of complex systems can be related to the properties of their underlying determinants," Proc. Natl. Acad. Sci. U.S.A. **76**, 5413–5417 (1979).
[23] B. B. Mandelbrot, *The Fractal Geometry of Nature* (Freeman, New York, 1983).
[24] M. Schroeder, *Fractals, Chaos, Power Laws* (Freeman, New York, 1991).
[25] L. Strauss, *Wave Generation and Shaping* (McGraw–Hill, New York, 1960).
[26] H. W. Bode, *Network Analysis and Feedback Amplifier Design* (Van Nostrand, Princeton, NJ, 1945).
[27] M. A. Savageau, *Biochemical Systems Analysis: A Study of Function and Design in Molecular Biology* (Addison-Wesley, Reading, MA, 1976).
[28] E. O. Voit, *Canonical Nonlinear Modeling: S-System Approach to Understanding Complexity* (Van Nostrand Reinhold, New York, 1991).
[29] N. V. Torres, E. O. Voit, and C. H. González-Alcón, "Optimization of nonlinear biotechnological processes with linear programming. Application to citric acid production in *Aspergillus niger*," Biotechnol. Bioeng. **49**, 247–258 (1996).
[30] T.-C. Ni and M. A. Savageau, "Model assessment and refinement using strategies from biochemical systems theory: Application to metabolism in human red blood cells," J. Theor. Biol. **179**, 329–368 (1996).
[31] D. C. Lewis, "A qualitative analysis of S-systems: Hopf bifurcations," in *Canonical Nonlinear Modeling. S-System Approach to Understanding Complexity*, edited by E. O. Voit (Van Nostrand Reinhold, New York, 1991), pp. 304–344.
[32] D. H. Irvine and M. A. Savageau, "Efficient solution of nonlinear ordinary differential equations expressed in S-system canonical form," SIAM (Soc. Ind. Appl. Math.) J. Numer. Anal. **27**, 704–735 (1990).
[33] S. A. Burns and A. Locascio, "A monomial-based method for solving systems of non-linear algebraic equations," Int. J. Numer. Methods Eng. **31**, 1295–1318 (1991).
[34] M. A. Savageau, "Finding multiple roots of nonlinear algebraic equations using S-system methodology," Appl. Math. Comput. **55**, 187–199 (1993).
[35] M. A. Savageau, "The behavior of intact biochemical control systems," Curr. Top. Cell. Regul. **6**, 63–130 (1972).
[36] M. A. Savageau, "A theory of alternative designs for biochemical control systems," Biomed. Biochim. Acta **44**, 875–880 (1985).
[37] R. Alves and M. A. Savageau, "Extending the method of mathematically controlled comparison to include numerical comparisons," Bioinformatics **16**, 786–798 (2000).
[38] M. A. Savageau, "Design of molecular control mechanisms and the demand for gene expression," Proc. Natl. Acad. Sci. U.S.A. **74**, 5647–5651 (1977).
[39] M. A. Savageau, "Models of gene function: general methods of kinetic analysis and specific ecological correlates," in *Foundations of Biochemical Engineering: Kinetics and Thermodynamics in Biological Systems*, edited by H. W. Blanch, E. T. Papoutsakis, and G. N. Stephanopoulos (American Chemical Society, Washington, DC, 1983), pp. 3–25.
[40] M. A. Savageau, "Regulation of differentiated cell-specific functions," Proc. Natl. Acad. Sci. U.S.A. **80**, 1411–1415 (1983).
[41] V. I. Arnol'd, *Catastrophe Theory* (Springer-Verlag, New York, 1992).
[42] A. Novick and M. Weiner, "Enzyme induction as an all-or-none phenomenon," Proc. Natl. Acad. Sci. U.S.A. **43**, 553–566 (1957).
[43] J. R. Sadler and A. Novick, "The properties of repressor and the kinetics of its action," J. Mol. Biol. **12**, 305–327 (1965).

Downloaded 20 Mar 2001 to 141.214.51.241. Redistribution subject to AIP copyright, see http://ojps.aip.org/chaos/chocr.jsp

[44] E. M. Cooke, *Escherichia coli and Man* (Churchill Livingstone, London, 1974).
[45] M. A. Savageau, "Demand theory of gene regulation: Quantitative development of the theory," Genetics **149**, 1665–1676 (1998).
[46] M. A. Savageau, "Demand theory of gene regulation: Quantitative application to the lactose and maltose operons of *Escherichia coli*," Genetics **149**, 1677–1691 (1998).
[47] J. H. Bond and M. D. Levitt, "Quantitative measurement of lactose absorption," Gastroenterology **70**, 1058–1062 (1976).
[48] J.-R. Malagelada, J. S. Robertson, M. L. Brown, M. Remington, J. A. Duenes, G. M. Thomforde, and P. W. Carryer, "Intestinal transit of solid and liquid components of a meal in health," Gastroenterology **87**, 1255–1263 (1984).
[49] J. H. Cummings and H. S. Wiggins, "Transit through the gut measured by analysis of a single stool," Gut **17**, 219–223 (1976).
[50] J. S. S. Gear, A. J. M. Brodribb, A. Ware, and J. T. Mann, "Fiber and bowel transit times," Br. J. Nutr. **45**, 77–82 (1980).
[51] M. A. Savageau, "*Escherichia coli* habitats, cell types, and molecular mechanisms of gene control," Am. Nat. **122**, 732–744 (1983).
[52] L. Hayflick, "The cellular basis for biological aging," in *Handbook of the Biology of Aging*, edited by C. E. Finch and L. Hayflick (Van Nostrand Reinhold, New York, 1977), pp. 159–186.
[53] H. I. Sears, I. Brownlee, and J. K. Uchiyama, "Persistence of individual strains of *E. coli* in the intestinal tract of man," J. Bacteriol. **59**, 293–301 (1950).
[54] H. I. Sears and I. Brownlee, "Further observations on the persistence of individual strains of *Escherichia coli* in the intestinal tract of man," J. Bacteriol. **63**, 47–57 (1952).
[55] D. A. Caugant, B. R. Levin, and R. K. Selander, "Genetic diversity and temporal variation in the *E. coli* population of a human host," Genetics **98**, 467–490 (1981).
[56] S. Garges, "Activation of transcription in *Escherichia coli:* The cyclic AMP receptor protein," in *Transcription: Mechanisms and Regulation*, edited by R. C. Conaway and J. W. Conaway (Raven, New York, 1994), pp. 343–352.

Downloaded 20 Mar 2001 to 141.214.51.241. Redistribution subject to AIP copyright, see http://ojps.aip.org/chaos/chocr.jsp

Biophysical Journal Volume 79 November 2000 2290–2304

Effect of Overall Feedback Inhibition in Unbranched Biosynthetic Pathways

Rui Alves*†‡ and Michael A. Savageau*

*Department of Microbiology and Immunology, University of Michigan Medical School, Ann Arbor, Michigan 48109-0620 USA, †Grupo de Bioquimica e Biologia Teoricas, Instituto Rocha Cabral, 1250 Lisboa, and ‡Programa Gulbenkian de Doutoramentos em Biologia e Medicina, Departamento de Ensino, Instituto Gulbenkian de Ciencia, 1800 Oeiras, Portugal

ABSTRACT We have determined the effects of control by overall feedback inhibition on the systemic behavior of unbranched metabolic pathways with an arbitrary pattern of other feedback inhibitions by using a recently developed numerical generalization of Mathematically Controlled Comparisons, a method for comparing the function of alternative molecular designs. This method allows the rigorous determination of the changes in systemic properties that can be exclusively attributed to overall feedback inhibition. Analytical results show that the unbranched pathway can achieve the same steady-state flux, concentrations, and logarithmic gains with respect to changes in substrate, with or without overall feedback inhibition. The analytical approach also shows that control by overall feedback inhibition amplifies the regulation of flux by the demand for end product while attenuating the sensitivity of the concentrations to the same demand. This approach does not provide a clear answer regarding the effect of overall feedback inhibition on the robustness, stability, and transient time of the pathway. However, the generalized numerical method we have used does clarify the answers to these questions. On average, an unbranched pathway with control by overall feedback inhibition is less sensitive to perturbations in the values of the parameters that define the system. The difference in robustness can range from a few percent to fifty percent or more, depending on the length of the pathway and on the metabolite one considers. On average, overall feedback inhibition decreases the stability margins by a minimal amount (typically less than 5%). Finally, and again on average, stable systems with overall feedback inhibition respond faster to fluctuations in the metabolite concentrations. Taken together, these results show that control by overall feedback inhibition confers several functional advantages upon unbranched pathways. These advantages provide a rationale for the prevalence of this control mechanism in unbranched metabolic pathways in vivo.

INTRODUCTION

Biochemical control systems have been studied for more than 45 years. The discovery of control by molecular feedback inhibition in biochemical pathways was initially made in unbranched biosynthetic pathways (Umbarger, 1956; Yates and Pardee, 1956). In these pathways, the most common pattern of control is inhibition of the initial reaction by the final product of the pathway (end-product inhibition or overall feedback inhibition).

There are several criteria for the functional effectiveness of control in such pathways that can be used to evaluate the biological significance of the overall feedback inhibition mechanism. A biochemical pathway should be robust, i.e., it should function reproducibly despite perturbations in the values of the parameters that define the structure of the system. The operating point (state) of the system should be stable so that the system returns to the steady state following small random fluctuations in the values of the dependent variables; if not, the system tends to be dysfunctional because spurious environmental fluctuations will lead to loss of the steady state. The flux through the pathway should be responsive to changes in the demand for the final product. This ensures that the amount of material flowing through the pathway is intimately coupled to the metabolic needs of the cell. Finally, the system should be temporally responsive to changes, because, otherwise, the system is unlikely to be competitive in rapidly changing environments. [A more extensive discussion of these criteria and their quantification can be found in Savageau (1976) and Hlavacek and Savageau (1997).]

There have been several studies focused on the effect of control by overall feedback inhibition on the stability of unbranched pathways. In general, the first enzyme of the pathway is considered to be allosteric, whereas the others are considered to be Michaelian (e.g., Goodwin, 1963; Morales and McKay, 1967; Walter, 1969a,b, 1970; Viniegra-Gonzalez, 1973; Hunding, 1974; Rapp, 1976; Dibrov et al., 1981). The stability of an unbranched pathway with overall feedback inhibition and enzymes confined to one of two spatial compartments with diffusion between compartments has been studied by Costalat and Burger (1996). They found that stability can be increased by this type of compartmentation. These studies considered pathways with no internal feedback inhibitions.

Several other patterns involving control by inhibitory feedback can, in principle, perform the same qualitative functions as overall feedback inhibition. One such pattern is, for example, a sequence of feedback inhibitions in which each intermediate inhibits the reaction that immediately precedes it (Koch, 1967). Other patterns of internal feedback inhibition can be found by searching either the litera-

Received for publication 17 April 2000 and in final form 9 August 2000.

Address reprint requests to Michael A. Savageau, Univ. of Michigan, Dept. Microbiology & Immunology, 5641 Medical Science II, Ann Arbor, MI 48109-0620. Tel.: 734-764-1466; Fax: 734-763-7163; E-mail: savageau@umich.edu.

© 2000 by the Biophysical Society

0006-3495/00/11/2290/15 $2.00

ture or some of the databases for metabolism that are burgeoning on the world wide web (e.g., KEEG: http://www.genome.ad.jp/kegg/; ECOCYC: http://ecocyc.PangeaSystems.com/ecocyc/server.html; PUMA: http://www.unix.mcs.anl.gov/compbio/PUMA/Production/puma_graphics.html; EMP: http://wit.mcs.anl.gov//EMP/). However, even when intermediate feedback inhibition patterns exist, control by overall feedback inhibition remains a prevalent theme in biosynthetic pathways.

Savageau (1972, 1974, 1975, 1976) studied the function of various patterns of feedback inhibition and explained the prevalence of control by overall feedback inhibition by using arguments based on selection. He assumed that the design of a pathway is selected to optimize certain systemic characteristics, and then compared those systemic characteristics in unbranched pathways with overall feedback inhibition to the same characteristics in pathways with alternative inhibitory feedback designs. He showed that the pathway with control by overall feedback inhibition is more robust, i.e., less sensitive to perturbations in parameter values than the pathway with many alternative designs (Savageau, 1974).

The stability of cases with control by internal feedback inhibitions has also been examined (e.g., Savageau, 1976; Thron, 1991a,b; Demin and Kholodenko, 1993). These authors found that systems with internal feedback inhibitions have larger stability margins than systems without these interactions. They also determined that, for systems without internal feedback inhibition, control by overall feedback inhibition decreases the stability margins of the pathway.

In this paper, we consider unbranched pathways with all possible patterns of internal feedback inhibitions (the "fully-wired" case) and use all of the criteria mentioned above to determine the biological significance of control by overall feedback inhibition in such pathways. We use a technique called Mathematically Controlled Comparison that was originally developed to determine irreducible qualitative differences in systemic behavior of models with alternative regulatory designs for the same network of reactions (Savageau, 1972, 1976; Irvine and Savageau, 1985). This qualitative technique requires the existence of closed-form solutions for the steady state. Such solutions can be obtained by using the local S-system representation to characterize the pathway of interest. Important functional constraints are introduced by equating relevant steady-state properties of the alternative systems being compared. The limitations of this technique have been overcome by a recently developed generalization that uses numerical methods to obtain results that are general in a statistical sense (Alves and Savageau, 2000a).

METHODS

Alternative models and key systemic properties

Consider the unbranched pathways depicted in Fig. 1. The independent variable X_{n+1} represents the cell's demand for the end product X_n. If the cell requires large amounts of X_n, then the value of X_{n+1} will be high; if small amounts of X_n are required, then the value of X_{n+1} will be low. The dynamic behavior of such systems can be described in principle by a set of ordinary differential equations. There is no generic representation of these

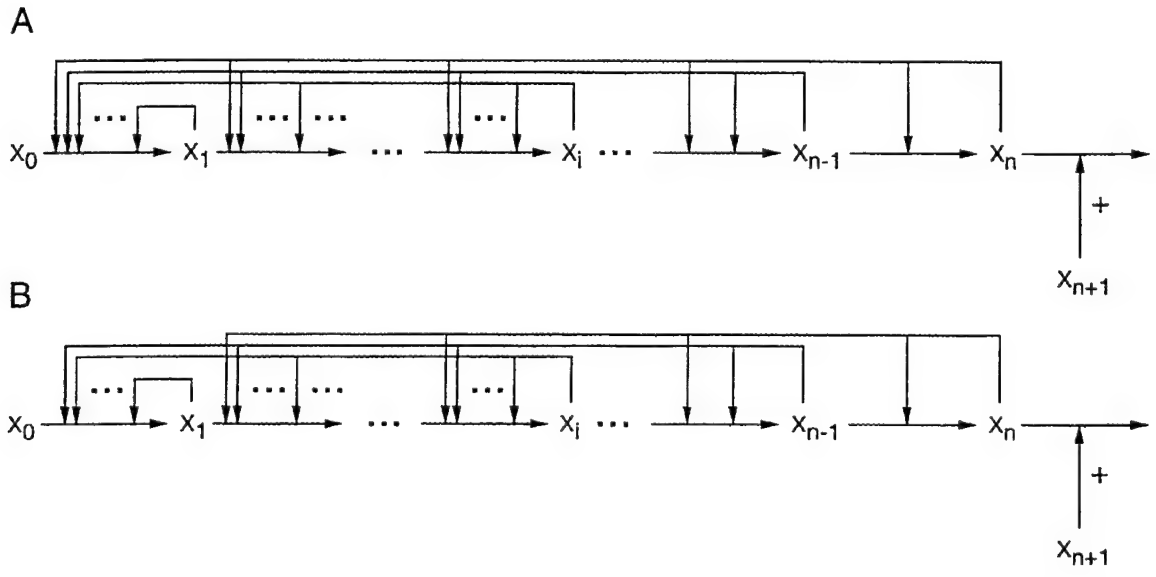


FIGURE 1 (*A*) Model of an unbranched pathway with all possible inhibitory feedback interactions (reference model). (*B*) Model of an unbranched pathway with all possible inhibitory feedback interactions except overall feedback inhibition (alternative model). The horizontal arrows represent biochemical reactions, whereas the vertical arrows represent inhibitory feedback interactions.

equations that can provide a globally accurate description of the behavior [see Appendix]. However, the set of equations can be approximated to the first order in logarithmic space (Savageau, 1969), yielding ordinary differential equations with the canonical form of an S-system (Savageau, 1996). This representation has a solid theoretical foundation based on Taylor's theorem. Thus, the validity of the results is guaranteed within some neighborhood of the nominal steady-state operating point. The size of this neighborhood cannot be specified in general, because it depends on the characteristics of each individual system.

For pathways with n intermediates, the general case in which all possible feedback inhibitions exist (Fig. 1 *A*) can be described in the local S-system representation as

$$\frac{dX_1}{dt} = \alpha_1 \prod_{j=0}^{n} X_j^{g_{1j}} - \alpha_2 \prod_{j=1}^{n} X_j^{g_{2j}}, \tag{1}$$

$$\frac{dX_i}{dt} = \alpha_i \prod_{j=i-1}^{n} X_j^{g_{ij}} - \alpha_{i+1} \prod_{j=i}^{n} X_j^{g_{i+1,j}}, \tag{2}$$

$$\frac{dX_n}{dt} = \alpha_n \prod_{j=n-1}^{n} X_j^{g_{nj}} - \alpha_{n+1} \prod_{i=n}^{n+1} X_j^{g_{n+1,j}}. \tag{3}$$

The corresponding case without overall feedback inhibition (Fig. 1 *B*) can be described by the same set of equations, except that Eq. 1 is replaced by

$$\frac{dX_1}{dt} = \alpha_1' \prod_{j=0}^{n-1} X_j^{g_{1j}'} - \alpha_2 \prod_{j=1}^{n} X_j^{g_{2j}}. \tag{4}$$

The rate law for each reaction is represented by a simple product of power-law functions. The values for the parameters in this representation can be determined directly from conventional experimental measurements of initial rate as a function of reactant and modifier concentrations (Savageau, 1976). The range of values for the concentrations is chosen to sample the region about the nominal steady state of interest.

The parameters are defined according to Taylor's theorem as

$$g_{ij} = \left(\frac{\partial \log v_i}{\partial \log X_j}\right)_0 = \left(\frac{\partial v_i}{\partial X_j}\right)_0 \left(\frac{X_{j0}}{v_{i0}}\right), \tag{5}$$

$$\alpha_i = v_{i0} \prod_{j=i-1}^{n} X_{j0}^{-g_{ij}}, \tag{6}$$

where the additional subscript zero signifies that the variables and their derivatives are evaluated at the steady-state operating point. The definition of these parameters allows them to be directly related to the parameters in other representations such as the traditional Michaelis–Menten representation. In the simplest case of the Hill rate law,

$$v = \frac{V_m X^n}{K_M^n + X^n} \tag{7}$$

[and the irreversible Michaelis–Menten rate law ($n = 1$)], these relationships are well known (Savageau, 1971a),

$$g = n \frac{K_M^n}{K_M^n + X_0^n}, \tag{8}$$

$$\alpha = v_0 X_0^{-g}. \tag{9}$$

The multiplicative parameters, α, can be interpreted as rate constants that are always positive. The exponential parameters, g, can be interpreted as kinetic orders that represent the direct influence of each intermediate on each rate law. If X_i is directly involved in the rate law V_j, either as a substrate or a modulator, and if an increase in X_i causes an increase in the rate V_j, then the kinetic order will be positive. If an increase in X_i causes a decrease in V_j, then the kinetic order will be negative. If X_i is not directly involved in V_j, then the kinetic order will be zero. The positive kinetic orders in Eqs. 1–4 are $g_{i+1,i}$ ($0 \le i \le n$) and g'_{10}, because these are the kinetic orders for substrates of reactions, and $g_{n+1,n+1}$, which, together with X_{n+1}, represents the demand for the end product X_n. The remaining kinetic orders, which represent feedback inhibitions, are negative.

At a steady state, the rate of production and the rate of consumption will be equal for each intermediate, and Eqs. 1–3 reduce to the following matrix equation (Savageau, 1969), which can be solved analytically.

$$\begin{bmatrix} b_1 - g_{10}Y_0 \\ b_2 \\ \vdots \\ b_{n-1} \\ b_n + g_{n+1,n+1}Y_{n+1} \end{bmatrix} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ a_{21} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots \\ a_{n-1,1} & \cdots & a_{n-1,n} \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_{n-1} \\ Y_n \end{bmatrix}, \tag{10}$$

where $b_1 = \log(\alpha_2/\alpha_1)$, $b_i = \log(\alpha_{i+1}/\alpha_i)$, $a_{ij} = g_{ij} - g_{i+1,j}$, and $Y_i = \log(X_i)$. Eq. 10 is linear and therefore easily solved to obtain the steady-state value for each Y_i, and then the corresponding value for each X_i is obtained by simple exponentiation. Eqs. 2–4 reduce to an identical matrix equation, except that the parameters of the first row are primed and $g'_{1n} = 0$.

Two types of systemic coefficients, logarithmic gains and parameter sensitivities, can be used to characterize the steady state of such models (Shiraishi and Savageau, 1992). Logarithmic gains measure the relative influence of each independent variable on each dependent variable of the integrated model. For example,

$$L(X_i, X_0) = \frac{d \log(X_i)}{d \log(X_0)} = \frac{dY_i}{dY_0} \tag{11}$$

measures the percent change in the concentration of intermediate X_i caused by a percentage change in the concentration of the initial substrate X_0. Logarithmic gains provide important information concerning the amplification or attenuation of signals as they are propagated through the system. The experimental measurement of a logarithmic gain involves the determination of steady-states fluxes and concentrations at different values for a given independent variable (Savageau, 1971a).

Parameter sensitivities measure the relative influence of each parameter on each dependent variable of the model. For example,

$$S(X_i, p_j) = \frac{d \log(X_i)}{d \log(p_j)} = p_j \frac{dY_i}{dp_j} \tag{12}$$

measures the percent change in the concentration of intermediate X_i caused by a percentage change in the value of the parameter p_j. Parameter sensitivities provide important information about system robustness, i.e., how sensitive the system is to perturbations in the parameters that define the structure of the system. Because enzymes usually have a first-order influence on the process they catalyze, the logarithmic gain in flux and in each concentration with respect to change in the concentration of each enzyme is the same as the sensitivity in flux and in each concentration with respect to change in the rate constant of the corresponding enzyme. The experimental measurement of a parameter sensitivity involves the determination of steady-state fluxes and concentrations before and after changing the value of a parameter by mutation or other means (Savageau, 1971b).

Because we can calculate closed-form steady-state solutions for Eqs. 1–3 and 2–4, we can also calculate each of the two types of coefficients

simply by taking the appropriate derivatives of those solutions. Although the mathematical operations involved are the same in each case, it is important to keep in mind that the biological significance of the two types of coefficients is very different.

The local stability of the steady state can be determined by applying the Routh criteria (Dorf, 1992). The magnitude of the two critical Routh conditions can be used to quantify the margin of stability (e.g., Savageau, 1976).

The use of the S-System formalism allows for an analytical study of the dynamical systems at steady state. Comparisons of systems with only one feedback inhibition to systems without feedback regulation can be done and interpreted in a fully symbolic way. However, for comparisons involving many feedback inhibitions, numerical values must be introduced for the parameters to make the comparisons interpretable. The steady-state behavior of the alternative models is compared with respect to their flux, intermediate concentrations, logarithmic gains with respect to changes in initial substrate and demand for end product, robustness, and stability margins. The differential equations also are solved numerically to characterize the temporal responsiveness of the alternative designs. To evaluate this, we increase the steady-state concentration of each X_i by 20% and measure the time the system takes to relax back to within 1% of its original steady state.

Calculating constraints for the mathematically controlled comparison

Only the first step in the pathway is allowed to differ between the reference model (Fig. 1 *A*) and the alternative model (Fig. 1 *B*). Therefore, to establish "internal equivalence" (Savageau, 1972, 1976; Irvine, 1991) between the two designs, we require the values for the corresponding parameters of all other steps in the two models to be the same.

The first step of the pathway differs between the reference model and the alternative model, and the degrees of freedom associated with this difference must be eliminated to the extent possible. If we reason that loss or gain of an inhibitory site on the first enzyme comes about by mutation, and that this mutation can cause changes in all the parameters of the first reaction, then a mutation causing loss of overall feedback inhibition would change the parameters α_1 and g_{10} through g_{1n} in Eq. 1 to the corresponding primed parameters in Eq. 4. Clearly, the value of the parameter g'_{1n}, which equals zero, differs from that of g_{1n}, which is nonzero. The remaining primed parameters also will have values that, in general, are not equal to the values of the corresponding parameters in the reference model. Because we wish to determine those effects that are due solely to changes in the structure of the system and not simply to arbitrary changes in the values of parameters, we shall specify values for the primed parameters that minimize all other effects. This can be accomplished by deriving the mathematical expression for a given steady-state property in each of the two models, equating these expressions, and then solving the constraint equation for the value of a primed parameter. This process establishes an "external equivalence" between the two designs (Savageau, 1972, 1976; Irvine, 1991). After values for all the primed parameters have been specified in terms of the known values for the reference system, the extra degrees of freedom have been eliminated, and we can proceed with the comparison.

Three classes of constraint equations are used to fix the values for the $k + 2$ primed parameters when there are k interactions that feed back to the first step of the alternative pathway. These are obtained by equating steady-state logarithmic gains, concentrations, and parameter sensitivities as described below.

First, equating the logarithmic gains for any one of the metabolites with respect to change in the initial substrate,

$$L(X_i, X_0)_A = L(X_i, X_0)_B \quad i = 1, 2, \ldots, n, \tag{13}$$

which causes each of the other corresponding intermediates to have the same logarithmic gain, specifies the value of the kinetic order g'_{10} in terms of known values for the reference system. This condition also makes the corresponding logarithmic gain in flux the same for the two designs.

Second, equating the concentrations for any one of the metabolites in the pathway,

$$Y_{iA} = Y_{iB} \quad i = 1, 2, \ldots, n, \tag{14}$$

which causes each of the corresponding intermediates to have the same concentration, specifies the value of the rate constant α'_1. This condition also makes the flux the same for the two designs.

Finally, the remaining $k - 1$ primed parameters are fixed by equating the rate-constant parameter sensitivities,

$$S(X_i, \alpha_j)_A = S(X_i, \alpha_j)_B \quad i = 1, 2, \ldots, n \quad j = 1, 2, \ldots, n, \tag{15}$$

for any X_i and $k - 1$ different rate constants α_j. Different results will be obtained, depending upon which of the parameter sensitivities are not used in this procedure.

For example, consider the case in which all $n - 1$ intermediates feed back on the first step in the pathway. If the unconstrained sensitivity in Eq. 15 is $S(X_n, \alpha_n)$, then the values of the primed parameters are given by

$$\log(\alpha'_1) = \log(\alpha_1) + \frac{g_{1n}}{g_{n+1,n} - g_{nn}} \log(\alpha_{n+1}/\alpha_n), \tag{16}$$

$$g'_{1p} = g_{1p} \quad 0 \le p < n - 1, \tag{17}$$

$$g'_{1,n-1} = g_{1,n-1} + \frac{g_{1n}}{g_{n+1,n} - g_{nn}} g_{n,n-1}. \tag{18}$$

If the unconstrained sensitivity in Eq. 15 is $S(X_i, \alpha_1)$, then the values of the primed parameters are

$$\log(\alpha'_1) = \frac{g_{n+1,n}}{g_{n+1,n} - g_{1n}} \log(\alpha_1) - \frac{g_{1n}}{g_{n+1,n} - g_{1n}} \log(\alpha_{n+1}), \tag{19}$$

$$g'_{1p} = \frac{g_{2n}}{g_{2n} - g_{1n}} g_{1p} \quad 0 \le p \le n - 1. \tag{20}$$

If the unconstrained sensitivity in Eq. 15 is $S(X_i, \alpha_j)$ where $1 < j < n$, then the values of the primed parameters are

$$\log(\alpha'_1) = \log(\alpha_1) - \frac{g_{1n}}{g_{jn} - g_{j+1,n}} \log(\alpha_{n+1}/\alpha_{j+1}), \tag{21}$$

$$g'_{1p} = g_{1p} \quad 0 \le p \le j - 1 \tag{22}$$

$$g'_{1p} = g_{1p} - \frac{g_{1n}}{g_{jn} - g_{j+1,n}} g_{2j} \quad j - 1 < p. \tag{23}$$

Because the objective of a controlled comparison is to minimize the differences between the systems being compared, we chose the unconstrained sensitivity that leads to the smallest number of systemic properties with values that differ between the reference system and the alternative system. The systemic differences are minimized when the unconstrained sensitivity is $S(X_i, \alpha_{n+1})$; any other choice leads to at least one additional systemic property that differs between the two systems.

If only a subset of the intermediates feed back on the first step of the pathway, and if we use the constraint set that causes the smallest number of properties to be different between systems A and B, then each kinetic order representing a feedback inhibition has the same value in both models, except for the kinetic order representing the last intermediate to feed back on the first step of the pathway. In general,

$$g'_{1k} = g_{1k} + \frac{g_{1n}}{\Delta_{nk}} \prod_{p=k}^{n-1} g_{p+1,p}, \tag{24}$$

where X_k is the last intermediate to feed back on the first step of the pathway, and Δ_{nk} is a positive subdeterminant of [A] that depends on the actual X_k and on the length n of the pathway. The kinetic orders g_{1p} with $p < k$ are the same for both systems. As for the rate constant α'_1, its general form is

$$\log(\alpha'_1) = \log(\alpha_1) + \sum_{p=k}^{n} \frac{\chi_{kp}}{\Delta_{np}} \log(\alpha_{n+1}/\alpha_{p+1}), \tag{25}$$

where χ_{kp} is either a function of the kinetic orders or zero.

For the special case in which the final product is the only metabolite to feed back on the initial step, the primed parameters are given by

$$\log(\alpha'_1) = \log(\alpha_1)\frac{g_{n+1,n}}{g_{n+1,n} - g_{1n}} - \log(\alpha_{n+1})\frac{g_{1n}}{g_{n+1,n} - g_{1n}}, \tag{26}$$

$$g'_{10} = \frac{g_{n+1,n}}{g_{n+1,n} - g_{1n}} g_{10}. \tag{27}$$

This means that g'_{10} is always smaller than g_{10}. (To contrast these results with the analogous results expressed within the Michaelis–Menten formalism, see the Appendix.)

Numerical comparison

It is straightforward to compare analytically corresponding magnitudes from each of the two designs. For two- and three-step pathways, the comparisons are clearly interpretable for most systemic properties. The analytical results give qualitative information that characterizes the role of overall feedback inhibition for the system in Fig. 1 *A*. As the length of the pathway increases, the analytical interpretation becomes problematic. To determine if a given magnitude is larger in the reference system or the alternative system requires knowledge of actual parameter values in these cases and a method, such as that found in Alves and Savageau (2000a), for making the numerical equivalent of a mathematically controlled comparison.

To obtain numerical information, one must introduce specific values for the parameters and compare systems. For this purpose, we have randomly generated a large ensemble of parameter sets and selected 5000 of these sets that define systems consistent with various physical and biochemical constraints. These constraints include mass balance, low concentrations of intermediates and small changes in their values to minimize utilization of the limited solvent capacity in the cell, small values for parameter sensitivities so as to desensitize the system to spurious fluctuations affecting its structure, and stability margins large enough to ensure local stability of the systems. A detailed description of these methods can be found in Alves and Savageau (2000c). Mathematica (Wolfram, 1997) was used for all the numerical procedures. Pathways of up to seven steps were studied using this numerical methodology.

To interpret the ratios that result from our analysis, we use density of ratios plots as defined in Alves and Savageau (2000b). The primary density plots of the raw data have the magnitude of some property for the reference system on the x-axis and the corresponding ratio of magnitudes (alternative system to reference system) on the y-axis. The primary plot can be viewed as a list of 5000 paired values that can be ordered with respect to the reference magnitude, thereby forming a list L_1 in which the first pair has the lowest measured value for property P in the reference model, the second has the second lowest, and so on.

Secondary density plots are constructed from the primary plots by the use of moving quantile techniques with a window size of 500. The procedure is as follows. One collects the first 500 ratios from the list L_1, calculates the quantile of interest for this sample, and pairs this number $\langle R \rangle$ with the median value of the corresponding P values for the reference model, denoted $\langle P \rangle$. One advances the window by one position, collects ratios 2–501, calculates $\langle R \rangle$, and pairs it with the corresponding $\langle P \rangle$ value and continues in this manner until the last ratio from the list L_1 was used for the first time (for further explanation of moving median techniques see, e.g., Hamilton, 1994).

The slope in the secondary plot measures the degree of correlation between the quantities plotted on the x- and y-axes. This technique also is used to examine correlations between ratios of interest and other magnitudes shared by the two systems, e.g., the correlation between the ratio of stability margins and the magnitude of a rate constant common to the two systems (for traditional applications of correlation analysis, see Wherry, 1984).

RESULTS

Mathematically controlled comparison

Response to availability of substrate and demand for end product

The responsiveness of each system to changes in the independent concentration variables X_0, which represents the availability of initial substrate, and X_{n+1}, which represents the demand for end product, is characterized by a set of logarithmic gains that provides a quantitative measure of signal propagation through the system.

The logarithmic gains of the two systems in response to changes in the initial substrate are identical at each step in the pathway [i.e., $L(V_i, X_0)_A = L(V_i, X_0)_B$ and $L(X_i, X_0)_A = L(X_i, X_0)_B$ for $1 \le i \le n$] because of the constraints for external equivalence described in the Methods section. Hence, the responsiveness of the two systems to changes in the availability of initial substrate is identical.

In contrast, the responsiveness of the two systems to changes in the demand for their end product is different. The ratio of the logarithmic gains in flux is given by

$$\left|\frac{L(V, X_{n+1})_A}{L(V, X_{n+1})_B}\right| = \left|1 + \frac{g_{1n}}{\zeta} \prod_{j=1}^{n} g_{j+1,j}\right| > 1, \tag{28}$$

where ζ is always a negative sum of products of the kinetic orders, $g_{1n} < 0$, and $g_{j+1,j} > 0$ for $j = 1, 2, \ldots, n-1$. These results demonstrate that the flux in the reference system is more responsive than that in the alternative system to changes in demand for end product.

The ratio of the logarithmic gains in concentration is given by

$$\left|\frac{L(X_i, X_{n+1})_A}{L(X_i, X_{n+1})_B}\right| = \left|1 + \frac{g_{1n}}{\zeta} \prod_{j=1}^{n-1} g_{j+1,j}\right| \quad i = 1, 2, \ldots, n, \tag{29}$$

where ζ is a sum of products of the kinetic orders that depends on i and the length of the pathway, $g_{1n} < 0$, and $g_{j+1,j} > 0$ for $j = 1, 2, \ldots, n-1$. When $i = 1$ or $i = n$, ζ is always positive and, thus, the reference model is always less sensitive to demand. When $1 < i < n$, ζ is positive in most cases. This shows that the concentrations are usually less sensitive to demand in the system with overall feedback inhibition.

Robustness of flux

The robustness of any systemic property with respect to perturbations in the values of the parameters that define the system is characterized by a set of parameter sensitivities. The steady-state flux of reference and alternative systems has different sensitivities with respect to the parameters α_n, α_{n+1}, $g_{1,n-1}$, $g_{n+1,n}$, g_{nn}, and $g_{n,n-1}$ that are common to the two systems. The sensitivities are the same with respect to all other parameters common to the two systems.

The sensitivities of the steady-state flux with respect to the parameters α_n, g_{nn}, and $g_{n,n-1}$ exhibit a common pattern. If we take the ratio of a sensitivity in the reference system to the corresponding sensitivity in the alternative system, we find that the ratio of the sensitivities is always less than 1. That is,

$$\left|\frac{S(V,p)_A}{S(V,p)_B}\right| = \left|1 + \frac{g_{1n}}{\gamma}\prod_{j=1}^{n-1} g_{j+1,j}\right| < 1, \qquad (30)$$

where γ is a positive sum of products of the kinetic orders, $g_{1n} < 0$, $g_{j+1,j} > 0$ for $j = 1, 2, \ldots, n-1$, and

$$-1 \le \frac{g_{1n}}{\gamma}\prod_{j=1}^{n-1} g_{j+1,j} < 0. \qquad (31)$$

Thus, the flux in the reference system is less sensitive to parameter variations, i.e., is more robust than that in the alternative system.

The sensitivities of the steady-state flux with respect to the parameter $g_{1,n-1}$ exhibit a similar pattern. The ratio of the sensitivities in this case is given by

$$\left|\frac{S(V,p)_A}{S(V,p)_B}\right| = \left|1 - \frac{g_{1n}g_{n,n-1}}{g_{1n}g_{n,n-1} - g_{1,n-1}(g_{nn} - g_{n+1,n})}\right| < 1. \qquad (32)$$

Although the function of the kinetic orders is different from that in Eq. 30, the flux in the reference system is again less sensitive to parameter variations, i.e., is more robust than that in the alternative system.

In contrast, the ratio of the sensitivities with respect to the parameters α_{n+1} and $g_{n+1,n}$ exhibits a different pattern,

$$\left|\frac{S(V,p)_A}{S(V,p)_B}\right| = \left|1 + \frac{g_{1n}}{\zeta}\prod_{j=1}^{n-1} g_{j+1,j}\right| > 1, \qquad (33)$$

where ζ is a negative sum of products of the kinetic orders. These parameter sensitivities are related to the last enzyme and reflect the design for responsiveness to changes in demand for end product.

As the position of the last intermediate that provides feedback inhibition to the first reaction approaches the beginning of the pathway, the number of sensitivities that differ between reference and alternative systems increases. This is so because the number of primed parameters decreases and a smaller number of conditions for external equivalence are needed to eliminated the extra degrees of freedom. In general, if the last intermediate that provides an inhibitory feedback to the first reaction is X_k for $k < n-1$, then the sensitivities of the flux to the rate constants α_k to α_{n+1} and those to the kinetic orders g_{ij} ($k \le i \le n$ and $i \le j \le n$) will differ between the reference and the alternative systems. In most cases, the sensitivities will be less in the reference system. There are exceptions to this, depending on the length of the pathway and on the last intermediate that provides feedback inhibition to the first step, and, in the case of α_{n+1} and $g_{n+1,n}$, the sensitivities of the reference system will always be greater, for the reasons we have already mentioned.

Robustness of concentrations

The steady-state concentrations of reference and alternative systems have different sensitivities with respect to many parameters that define the systems. In some cases, the ratio of the corresponding sensitivities is always <1 or always >1, but, in others, the ratio is <1 for some values of the parameters and >1 for other values. In the latter cases, an examination of actual numerical values for the parameters is critical.

The ratio of sensitivities for the concentration of each intermediate in the pathway with respect to changes in the kinetic order $g_{1,n-1}$ is identical to that given in Eq. 32. Similarly, the ratio of sensitivities for X_n with respect to changes in the rate constants α_n or α_{n+1} is always of the form

$$\left|\frac{S(X_n,\alpha_p)_A}{S(X_n,\alpha_p)_B}\right| = \left|1 + \frac{g_{1n}}{\zeta_p}\prod_{j=1}^{n-1} g_{j+1,j}\right| < 1 \quad p = n, n+1, \qquad (34)$$

where ζ_p is a different positive sum of products of kinetic orders for each α_p, $p = n, n+1$, and

$$-1 \le \frac{g_{1n}}{\zeta_p}\prod_{j=1}^{n-1} g_{j+1,j} < 0 \quad p = n, n+1. \qquad (35)$$

Thus, the reference system is always less sensitive to changes in these parameters.

In contrast, the ratio of sensitivities for X_n with respect to changes in the kinetic orders $g_{n+1,n}$, $g_{n,n-1}$, or g_{nn} is always of the form

$$\left|\frac{S(X_n, g_{pq})_A}{S(X_n, g_{pq})_B}\right| = \left|1 + \frac{g_{1n}}{\zeta_{pq}} \prod_{j=1}^{n-1} g_{j+1,j}\right|, \tag{36}$$

where ζ_{pq} is a different positive sum of products of the kinetic orders for each g_{pq}. In this case, the ratio can be >1 or <1. This means that the sensitivity of the reference system will be greater than the sensitivity of the alternative system for some values of the parameters and less for others. Similarly, the ratio of sensitivities for each intermediate X_i with respect to changes in each parameter can be $\gtrless 1$, depending on values of the parameters.

Again, as the position of the last intermediate that provides feedback inhibition to the first reaction approaches the beginning of the pathway, the number of sensitivities that differ between reference and alternative systems increases. In general, if the last intermediate that provides an inhibitory feedback to the first reaction is X_k, then the ratio of sensitivities for each metabolite with respect to changes in the kinetic order g_{1k} is given by

$$\left|\frac{S(X_i, g_{1k})_A}{S(X_i, g_{1k})_B}\right| = \left|1 + \frac{g_{1n}}{\zeta_{1k}} \prod_{j=1}^{n-1} g_{j+1,j}\right| < 1 \quad i = 1, 2, \ldots, n. \tag{37}$$

In this equation, ζ_{1k} is a positive subdeterminant of the [A] matrix. The ratio of sensitivities for the end product with respect to changes in each of the parameters common to the two systems also is always ≤ 1. Similarly, the ratio of sensitivities for the last intermediate that feeds back to the first reaction, X_k, with respect to the parameters α_k or g_{kj} ($k \leq j \leq n$) is always <1. Thus, the reference system is always more robust than the alternative system in these cases. As for the remaining cases, the sensitivities of the reference system will be greater than the sensitivities of the alternative system for some values of the parameters and less for others.

Stability

The characteristic equation for Eqs. 1–3 operating near the steady state can be written as

$$\begin{vmatrix} F_1a_{11} - \lambda & F_1a_{12} & & \cdots & F_1a_{1n} \\ F_2a_{21} & F_2a_{22} - \lambda & & \cdots & F_2a_{2n} \\ 0 & F_3a_{32} & & \cdots & F_3a_{3n} \\ \vdots & & \ddots & & \vdots \\ 0 & \cdots & F_{n-1}a_{n-1,n-2} & F_{n-1}a_{n-1,n-1} - \lambda & F_{n-1}a_{n-1,n} \\ 0 & \cdots & 0 & F_na_{n,n-1} & F_na_{nn} - \lambda \end{vmatrix} = 0, \tag{38}$$

where $F_i = V_{i0}/X_{i0}$ and $a_{ij} = g_{ij} - g_{i+1,j}$. Eq. 38 can be expanded into polynomial form and the Routh conditions for local stability determined. The last two Routh conditions are critical for stability (Frazer and Duncan, 1929). The last condition is equivalent to the condition $(-1)^n\det(A) > 0$, which is always true for the systems we are considering (Savageau, 1976, Appendix B).

The two critical Routh conditions for a two-step pathway are

$$R_1 = F_1(g_{11} - g_{21}) + F_2(g_{22} - g_{32}) < 0 \tag{39}$$

and

$$R_2 = F_1F_2[g_{11}(g_{22} - g_{32}) + g_{21}(g_{32} - g_{12})] > 0. \tag{40}$$

Both these conditions are always satisfied for both system A ($g_{12} < 0$) and system B ($g'_{12} = 0$ and $g'_{11} = g_{11} + g_{12}g_{21}/(g_{32} - g_{22}) < g_{11} < 0$), so these systems are always stable. The ratio of the last Routh condition for the two systems is equal to unity, whereas that for the penultimate condition is given by

$$\frac{R_{1A}}{R_{1B}} = 1 - \frac{F_1g_{12}g_{21}}{(F_1g_{12}g_{21} - F_1g_{11}g_{22} + F_1g_{21}g_{22} - F_2g_{22}^2 + F_1g_{11}g_{32} - F_1g_{21}g_{32} + 2F_2g_{22}g_{32} - F_2g_{32}^2)} < 1. \tag{41}$$

Thus, the stability margin is larger for the alternative system B.

The two critical Routh conditions for a three-step pathway are already considerably more complex. Whereas the last condition is always positive, the most critical condition is the penultimate one that can be positive or negative, depending upon the particular values for the parameters. The ratio of the last condition for the two systems is equal to 1; the ratio of the penultimate condition can be >1 or <1, depending on the values for the parameters. These same conclusions are obtained for pathways of length four or greater: the ratios cannot be determined analytically to be >1 or <1, and we must resort to numerical methods.

Transient time

There is no analytical way to accurately calculate the transient times of the pathway. This must be done numerically.

Numerical comparisons

Unlike the symbolic analysis performed in the previous section, using actual numbers for the values of the parameters limits the absolute generality of the results. However, it does allow us to obtain general conclusions in a statistical sense. The results described below have been obtained for pathways of up to seven intermediates. The trends in these results remain constant throughout all the tested lengths (i.e., pathways from 2 to 7 intermediates), which suggests that they will remain so for longer pathways. The use of these numerical methods allows us not only to study the

effects of overall feedback inhibition, but also to study correlations that exist between systemic properties and the different parameters of the system.

Response to availability of substrate and demand for end product

The logarithmic gains in concentrations of the two systems in response to changes in the initial substrate X_0 are identical at each step in the pathway because of the constraints for external equivalence described in the Methods section. The same is true for the logarithmic gains in flux. Hence, the responsiveness of concentrations and fluxes in the two systems to changes in the availability of initial substrate is identical numerically as well as analytically.

The logarithmic gain in flux for system A in response to changes in the demand for end product was shown analytically to be greater than that for system B. The graph of $L(V, X_{n+1})_A/L(V, X_{n+1})_B$ versus $L(V, X_{n+1})_A$ (Fig. 2 *A*), which is the moving median density of ratios plot introduced in Alves and Savageau (2000c), shows how much greater, on average, the response is for system A. It also shows a negative correlation between the ratio of responses and the response of the reference system. This means that, as $L(V, X_{n+1})_A$ increases, the ratio $L(V, X_{n+1})_A/L(V, X_{n+1})_B$ tends to decrease.

The logarithmic gain in end-product concentration for system A in response to changes in the demand for end product also was shown analytically to be smaller than that for system B. The graph of $L(X_n, X_{n+1})_A/L(X_n, X_{n+1})_B$ versus $L(X_n, X_{n+1})_A$ (Fig. 2 *B*) shows how much smaller, on average, the response is for system A. It also shows a positive correlation between the ratio of responses and the response of the reference system.

Robustness

Figure 2 shows typical moving median density of ratios plots for the aggregate parameter sensitivities of flux and concentrations. The aggregate parameter sensitivity of the flux V is smaller, on average, for system A (Fig. 2 *C*). Assume that X_k is the last intermediate to feed back on the first reaction of the pathway. The aggregate parameter sensitivity of X_k is smaller, on average, for system B (Fig. 2 *D*). The average difference in aggregate sensitivities for this metabolite is never larger than a few percent. With regard to the remaining intermediates, the graphs for X_i (Fig. 2 *E*) and X_j (Fig. 2 *F*) represent typical plots of aggregate parameter sensitivities. In these cases, we find that random reference systems are less sensitive than the equivalent alternative systems. The average differences can range from a few percent to fifty or more percent. The individual parameter sensitivities of X_n were analytically determined to be smaller in system A. In the example presented here, the difference is, on average, just a few percent (Fig. 2 *G*); however, depending on the length of the pathway, this difference can increase to more significant values.

The flux (Fig. 2 *C*) and concentrations X_i, $i < n$, (Fig. 2, *D*, *E*, and *F*) show a positive correlation between the ratio of their aggregate sensitivities in the two systems and the aggregate sensitivity in the reference system when its value is low. For systems with low sensitivities, system A is, on average, much less sensitive than system B. For higher values of the aggregate sensitivities in the reference system, there is no correlation. In the case of X_k, the ratio is fairly independent of the values of the aggregate sensitivity in the reference system.

Stability

The last critical Routh criterion is always the same in the reference and alternative systems, as has been shown analytically. For a two-step pathway, the margin of stability determined by the penultimate criterion is always larger in system B. For longer pathways, the margin of stability can be larger in either the reference or the alternative system, depending on the numerical values of the parameters. The differences between the two systems with respect to this penultimate criterion are small (on average less than 2%, Fig. 2 *H*), which implies that systems with and without overall feedback inhibition will have comparable stability margins.

Transient time

Fig. 2 *I* shows a typical moving median density of ratio plot for transient time. This plot shows that the reference system usually responds to perturbations in the steady state more quickly than the alternative system. For reference systems with a fast response to changes, the transient times can be, on average, half that of the corresponding alternative systems. For reference systems that are sluggish, the difference is, on average, smaller, though it still exists.

Effects of parameter values on systemic properties

Rate-constant effects on aggregate sensitivities

Assume that X_k is the last intermediate to feed back on the first reaction. Plotting the aggregate sensitivities as a function of α_j, $n \leq j$, shows that there is a correlation between each rate constant α_j and each of the aggregate sensitivities (Fig. 3 *A*). For small α_j, the correlation is either nonexistent or slightly negative, whereas, for large values, this correlation is positive. As for the other rate constants, with $j < n$, there are no obvious correlations that are general for all the pathway lengths studied, although, for some lengths, specific correlations are observed.

Kinetic-order effects on aggregate sensitivities

For X_n, the aggregate sensitivity is correlated with several parameters. There is a positive correlation between this

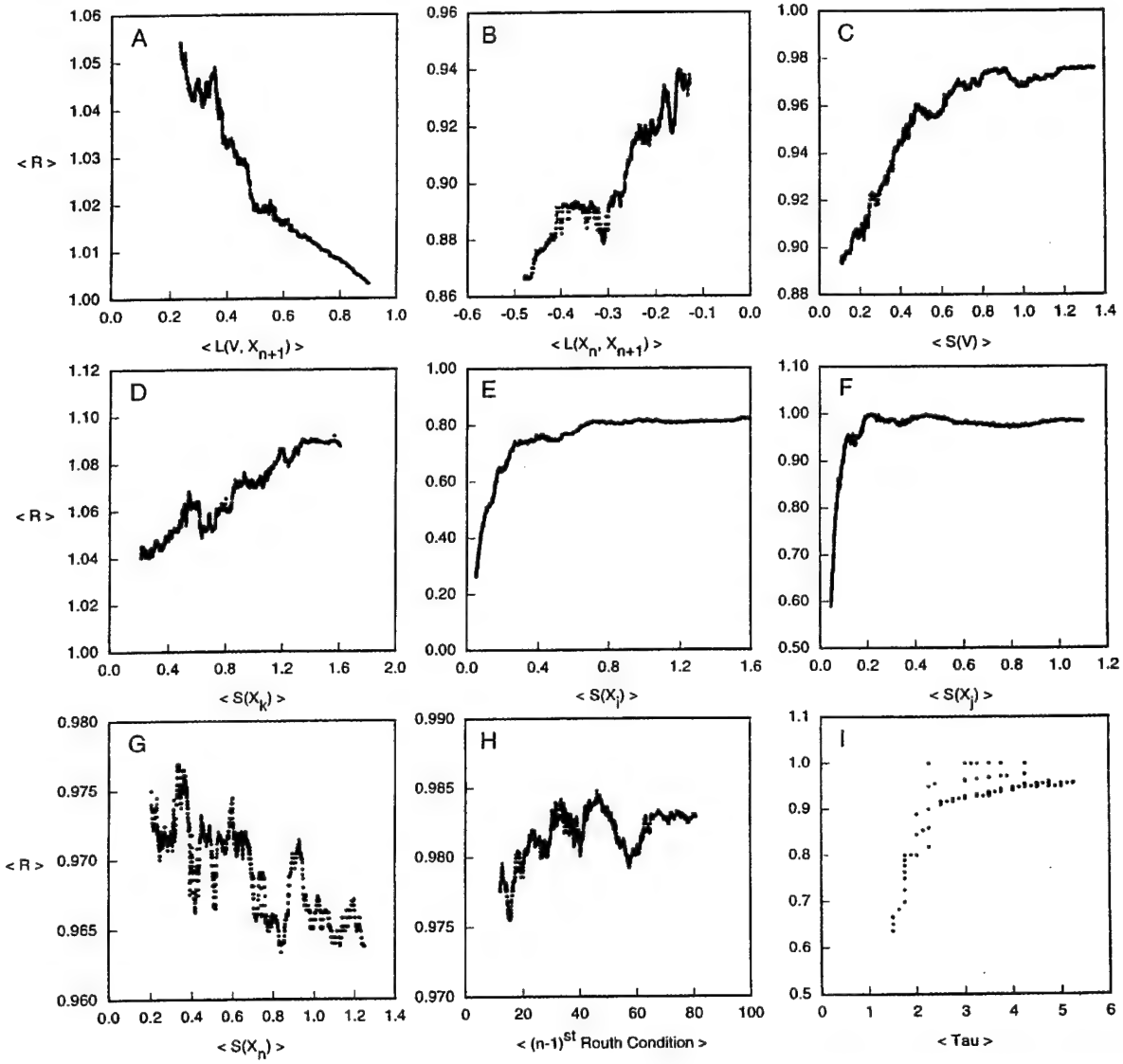


FIGURE 2 Typical moving median density of ratios plots for different magnitudes. The values on the *X*-axis represent the moving median of the relevant magnitude in the reference system. The values on the *Y*-axis represent the moving median of the ratio of that magnitude in the reference system to the corresponding magnitude in the alternative system. (*A*) Logarithmic gain in flux in response to changes in demand for the end product, $L(V, X_{n+1})$. (*B*) Logarithmic gain in end-product concentration in response to changes in demand for the end product, $L(X_n, X_{n+1})$. (*C*) Aggregate sensitivity of the pathway flux, $S(V)$. (*D*) Aggregate sensitivity of the concentration of the last intermediate to feed back on the first reaction, $S(X_k)$. (*E*) Aggregate sensitivity of the concentration of any intermediate in the pathway before X_k, $S(X_i)$. (*F*) Aggregate sensitivity of the concentration of any intermediate in the pathway after X_k, $S(X_j)$. (*G*) Aggregate sensitivity of the concentration of the end-product, $S(X_n)$. (*H*) The penultimate (i.e., $n - 1$st) Routh criterion; this represents the margin of stability. (*I*) Transient time, τ in normalized units, is the time the pathway takes to return within 1% of its steady state following a 15% perturbation in the steady-state values. Each of these plots is for a specific pathway length; only the parameter values are changed randomly. However, because the trends observed for different pathway lengths are the same, we have only shown a representative case.

sensitivity and g_{1n}. Because g_{1n} is always negative, this means that the aggregate sensitivity of X_n, $S(X_n)$, is usually smaller for high values of overall feedback inhibition. The same is true for the correlation between $S(X_n)$ and g_{in} when $i < n$ (Fig. 3 *B*). If $i = n$, there is a negative correlation between this aggregate sensitivity and g_{1n}. The correlation of the aggregate sensitivities of the other intermediates with g_{1n} is usually small or nonexistent. There is a negative correlation between the aggregate sensitivity of X_i and $g_{i+1,i}$ or $g_{n,n-1}$ (Fig. 3 *C*) and a positive correlation between that

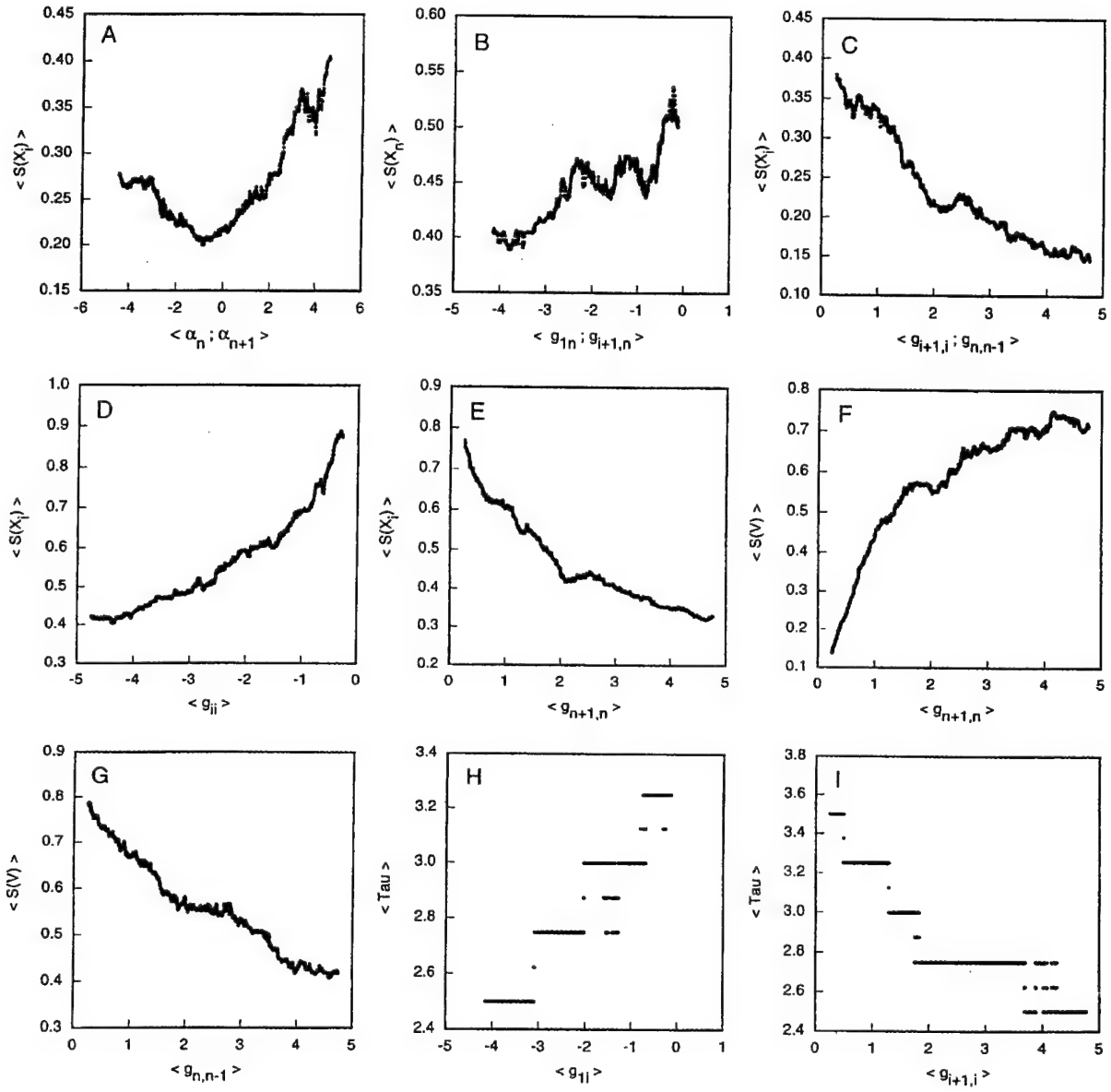


FIGURE 3 Typical moving median correlation plots between different systemic properties and different kinetic parameters of the reference system. The values on the *X*-axis represent the moving median of the relevant kinetic parameter. The values on the *Y*-axis represent the moving median of the relevant systemic property. (*A*) Aggregate sensitivity of the concentration of any pathway intermediate X_i versus the rate-constant parameters α_n or α_{n+1}. (*B*) Aggregate sensitivity of the concentration of the end product X_n versus the kinetic-order parameters g_{1n} or $g_{i+1,n}$. (*C*) Aggregate sensitivity of the concentration of any pathway intermediate X_i versus the kinetic-order parameters $g_{i+1,i}$ or $g_{n,n-1}$. (*D*) Aggregate sensitivity of the concentration of any pathway intermediate X_i versus the kinetic-order parameter g_{ii}. (*E*) Aggregate sensitivity of the concentration of the end product X_n versus the kinetic-order parameter $g_{n+1,n}$. (*F*) Aggregate sensitivity of the pathway flux V versus the kinetic-order parameter $g_{n+1,n}$. (*G*) Aggregate sensitivity of the pathway flux V versus the kinetic-order parameter $g_{n,n-1}$. (*H*) Transient time τ versus the kinetic-order parameter g_{1i}. (*I*) Transient time τ versus the kinetic-order parameter $g_{i+1,i}$. Each of these plots is for a specific pathway length; only the parameter values are changed randomly. However, because the trends observed for different pathway lengths are the same, we have only shown a representative case.

of X_i and g_{ii} (Fig. 3 *D*). Also, the aggregate sensitivity of each X is negatively correlated with $g_{n+1,n}$ (Fig. 3 *E*). These are the correlations that are generally observed for the aggregate sensitivities of concentrations, although other individual correlations can be found for specific intermediates and specific pathway lengths.

The correlations between aggregate sensitivities of flux and the various kinetic-order parameters are less clear. The

correlation with $g_{n+1,n}$ is positive for low values of $g_{n+1,n}$, but it disappears as the value of $g_{n+1,n}$ increases (Fig. 3 *F*). The only other general correlation observed is that between the aggregate sensitivity of the flux and the kinetic order $g_{n,n-1}$. This is a negative correlation that also vanishes as the value of $g_{n,n-1}$ increases. This can be seen in Fig. 3 *G*.

Rate-constant and kinetic-order effects on margin of stability

The correlations between a given Routh criterion and the various parameters depends on which criterion is considered. The results are pathway length-specific, and no general trend can be found.

Rate-constant and kinetic-order effects on transient time

There is no clear correlation between transient time and the various rate constants. There are, however, positive correlations between transient time and the kinetic orders g_{1i}, for $i \geq 1$ (Fig. 3 *H*). There also are negative correlations between transient time and the kinetic orders $g_{i+1,i}$ for $i > 1$ (Fig. 3 *I*). These were the only observed correlations with transient time.

Effects of enzyme levels on systemic variables

We have determined the logarithmic gains in flux and concentrations in response to changes in the level of individual enzymes. When comparing logarithmic gains in flux and concentrations in the reference and alternative systems, the equivalence conditions will make all corresponding coefficients identical except the last two. We also have examined the correlations among the logarithmic gains.

The last two logarithmic gains in concentrations are, on average, lower in the system controlled by overall feedback inhibition (see also Eq. 34). However, there is no general pattern of correlation among the logarithmic gains in concentrations.

The penultimate logarithmic gain in flux is always larger in the alternative system (Fig. 4 *C*). The last logarithmic gains in flux, which is a measure of coupling between flux and the demand for final product, is always larger in the reference system (Fig. 4 *D*). The logarithmic gains in flux with respect to changes in each individual enzyme except the last are directly correlated (Fig. 4 *A*, *B*, and *C*). The last logarithmic gain in flux is inversely correlated with all the others (Fig. 4 *D*). This is a well-known effect of feedback inhibition, i.e., it decreases the sensitivity of the flux through the system to parameters (in this case enzyme levels) inside the feedback loop while increasing the sensitivity to parameters outside the loop.

DISCUSSION

In this paper, we are addressing a generic property characteristic of an entire class of biochemical systems: Why is the pattern of overall feedback inhibition in unbranched biosynthetic pathways so prevalent? Because there are innumerable specific cases that could be examined, most of which have never arisen or may no longer exist because of natural selection, one could never hope to answer this type of question with an experimental approach. However, on a more fundamental level (beyond the sheer number of possibilities that would have to be constructed and examined), one must face the difficulty of performing even a single experimental comparison under well-controlled conditions so that the results will not be confused by extraneous differences.

The method of mathematically controlled comparison was developed specifically to address these issues. It allows one to examine enormous numbers of alternatives in parallel, more than would ever be possible by experimental means; it also allows essentially ideal controlled comparisons, comparisons that could only be done with an enormous experimental effort. In short, this is the type of question that is more appropriately answered by means of a theoretical analysis than by the accumulation of experimental evidence for one specific system after another.

The experimental difficulty in doing the equivalent of a mathematically controlled comparison can be seen from the expressions in the Appendix. One would first have to generate a large number of feedback-resistant mutants. Each independent mutant would, in general, have different values for the resulting K'_M and V'_m parameters. One would have to measure the K'_M for each of the mutants until one was found that had the appropriate value, as determined by the constraints for external equivalence in Eqs. A4–A8. If one was lucky enough to find that this mutant also had the correct value for V'_m, as determined by the constraints for external equivalence in Eqs. A4–A8, then one could measure the systemic differences between the wild-type and mutant to experimentally verify the theoretical results. If the V'_m value was not appropriate, one might construct a mutant strain with the structural gene for the first enzyme under the control of a promoter whose activity can be independently varied. In such a construct, one might be able to adjust the promoter activity to provide the appropriate value for V'_m. Again, one could measure the systemic differences between the wild-type and mutant to experimentally verify the theoretical results. As can be seen from this discussion of what it would take to do the experiments properly, it is unlikely that anyone would undertake the task. This is especially so when the result will only be valid for one special system, and will not contribute significantly to the validation of the general principle.

This discussion is in no way a criticism of the experimental approach. It simply acknowledges the fact that only

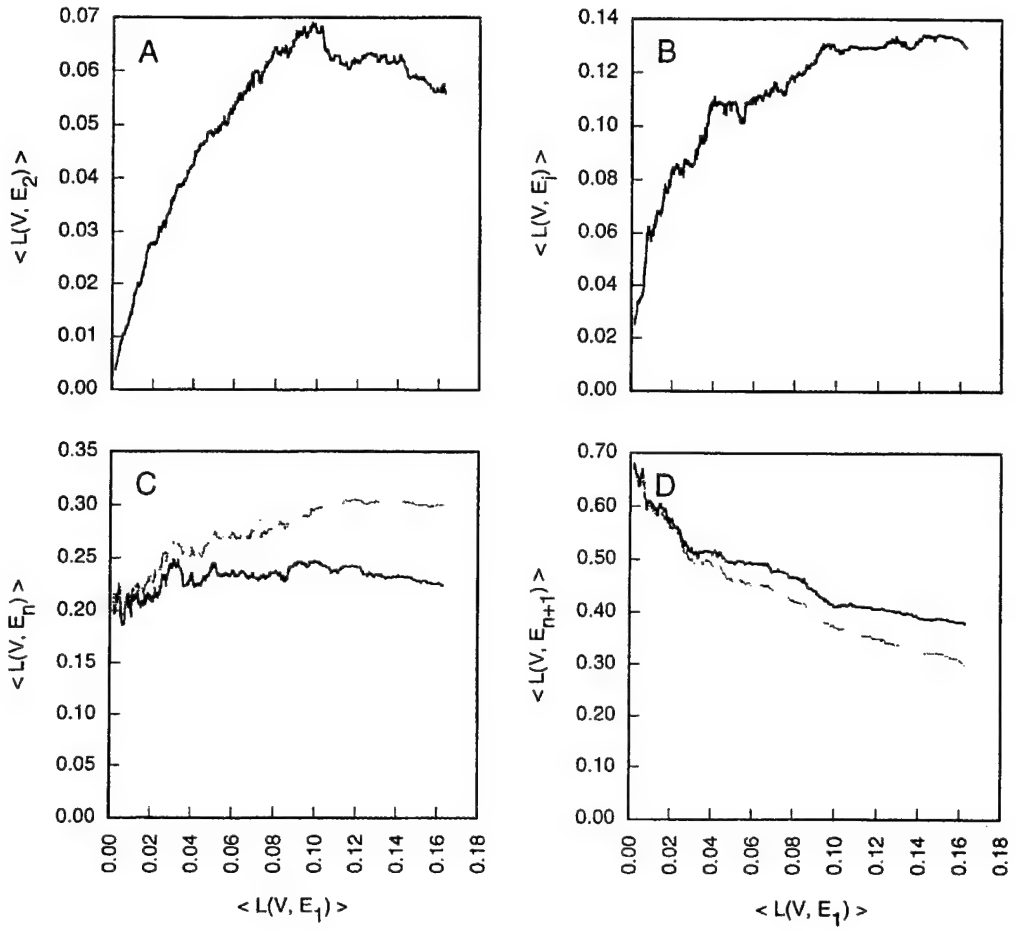


FIGURE 4 Typical moving median correlation plots between different logarithmic gains in flux with respect to changes in individual enzyme levels. The values on the *X*-axis represent the moving median of the logarithmic gain with respect to the first enzyme of a pathway. The values on the *Y*-axis represent the moving median of the logarithmic gains with respect to subsequent enzymes in the pathway. Full lines indicate curves for the reference system, and dashed lines indicate curves for the alternative system. (*A*) Logarithmic gain in flux with respect to the second enzyme of the pathway versus logarithmic gain in flux with respect to the first enzyme of the pathway. (*B*) Logarithmic gain in flux with respect to the *i*th enzyme of the pathway ($i \neq 1, 2, n, n+1$) versus logarithmic gain in flux with respect to the first enzyme of the pathway. (*C*) Logarithmic gain in flux with respect to the penultimate enzyme of the pathway versus logarithmic gain in flux with respect to the first enzyme of the pathway. (*D*) Logarithmic gain in flux with respect to the last enzyme of the pathway versus logarithmic gain in flux with respect to the first enzyme of the pathway. Each of these plots is for a specific pathway length; only the parameter values are changed randomly. However, because the trends observed for different pathway lengths are the same, we have only shown a representative case.

specific theoretical predictions are amenable to direct experimental test. More general theoretical predictions that apply to an entire class of systems require experimental information for many members of the class. The experimental validation of the theory presented here is the fact that it can account for the prevalence of overall feedback inhibition in biosynthetic pathways.

In this work, we have used a numerical generalization of the method of mathematical controlled comparison to examine systemic properties of models with and without overall feedback inhibition in unbranched pathways that otherwise have an arbitrary pattern of feedback inhibitions. In summarizing our findings, we shall interlace the results of the older analytical approach with those of the more recently developed numerical approach. This has the advantage of showing how the numerical approach goes beyond the analytical approach to broaden the scope of mathematical controlled comparison.

By using mathematically controlled comparisons, we have ensured that the systems achieve the same steady-state flux, metabolite concentrations, and logarithmic gains with respect to changes in the concentration of initial substrate, whether overall feedback inhibition is present or not. However, the alternative designs exhibit differences for many

other systemic properties. In the following seven types of results, the analytical approach yields unambiguous qualitative differences.

1. The logarithmic gain in flux resulting from an increase in demand for end product is always greater in the system with overall feedback inhibition. This ensures a tighter control of the material flowing through the pathway by the demand for such material.
2. The logarithmic gain in the concentration of the first and last metabolite resulting from an increase in demand for end product is always less in the system with overall feedback inhibition. This shows that these concentrations tend to be buffered against changes in demand for end product.
3. The sensitivities of the flux to changes in the parameters of the intermediate reactions for the system with overall feedback inhibition are less than or equal to those of the otherwise equivalent system without this inhibition. This shows that overall feedback inhibition increases the robustness of the flux.
4. The sensitivities of the flux to changes in the parameters of the last reaction for the system with overall feedback inhibition are greater than or equal to those of the otherwise equivalent system without this inhibition. This is related to the first point above.
5. The sensitivity of the end-product concentration to each rate-constant parameter of the system with overall feedback inhibition is always less than or equal to that of the otherwise equivalent system without this mechanism. This was shown to be analytically true independent of pathway length. The reference system is thus more effective in buffering the final product of the pathway against parameter fluctuations.
6. The sensitivity of each concentration to the parameter representing the last intermediate to feed back on the first reaction is always less in the system with overall feedback inhibition. Again, the reference system is better protected against fluctuations of this parameter.
7. For the special case of pathways with two intermediates, the alternative system has larger stability margins than the reference system with overall feedback inhibition. The more general case is discussed below.

From the above results, we conclude that pathway flux is more responsive to change in demand for the end product when overall feedback inhibition is present and that the concentration of final product, and the magnitude of pathway flux, is less sensitive to changes in the parameters of the system with overall feedback inhibition.

In each of the above results, the numerical method not only confirmed the qualitative differences, but also showed how large the differences were on average. In the following four types of results the analytical approach yields either no results or ambiguous qualitative differences, whereas the numerical approach gives statistical regularities in either situation.

1. The logarithmic gain in the concentration of intermediates X_2 to X_{n-1} resulting from an increase in demand for end product may be either larger or smaller in the reference system depending on the intermediate, the pathway length, or the values of the parameters. The numerical results show that, on average, these logarithmic gains are smaller in the reference system.
2. For all concentrations, there are some sensitivities that may be either larger or smaller in the reference system. The numerical approach shows that, on average, these concentrations have smaller aggregate sensitivities in the reference system. The differences between the reference system and the alternative system can range anywhere between a few percent to fifty percent or more, depending on the length of the pathway and the concentration of interest.
3. The stability margins for pathways longer than two reactions can be larger in either the reference system or the alternative system, depending on the values of the parameters. Use of the statistical methodology shows that, on average, overall feedback inhibition decreases the margin of stability. However, the differences between systems with and without overall feedback inhibition are, on average, less than 3% and typically less than 5%.
4. The transient time of the pathways cannot be determined analytically. Numerical results show that transient times tend to be smaller in pathways with overall feedback inhibition. Although a small percentage of systems with overall feedback inhibition have higher transient times, on average, overall feedback inhibition decreases transient times in stable systems. Systems with overall feedback inhibition can be, on average, a few percent faster to twice as fast as systems without overall feedback inhibition, depending on the length of the pathway.

In addition to resolving ambiguities in the analytical comparisons, the numerical methods allowed us to identify some general effects of parameter values on systemic properties. We found that there is a correlation between the values of α_j ($j = n, n + 1$) and the values of the aggregate sensitivities for each metabolite as well as the flux. For very low values of α_j, the aggregate sensitivities will not be strongly affected by a change in those parameters. As these parameters becomes larger than 1, a correlation develops. As the value of α_j increases, so does the aggregate sensitivity on average. The rate constant α_{n+1} is a parameter that can be interpreted as the demand for X_n. This means that, as the demand increases, so do the aggregated sensitivities. Why this happens is not clear.

General correlations between systemic properties and kinetic-order parameters also were identified. For example, we found that the transient times of the pathway are inversely correlated with the kinetic orders $g_{i+1,i}$. This means

that, on average, a system will respond faster to perturbations if the kinetic orders for the substrates of the reactions are higher. The perturbations that were given to the systems were always positive, i.e., the substrates were increased above their nominal steady-state values. Higher kinetic orders with respect to substrate mean that the rate will have a sharper response to an increase in the substrate, thus causing it to return to the steady-state value faster. In addition to this, there is a positive correlation between transient times and feedback parameters. Lower magnitudes for the kinetic orders representing inhibitory feedback make the rate less sensitive to increases in the concentrations of its inhibitors. Thus, after an increase in inhibitor concentrations, systems with lower magnitudes for the feedback interaction will have faster rates than systems with high magnitudes. It is not clear why these correlations exist only with respect to the parameters representing feedback to the first reaction of the pathway.

In conclusion, it is important to note that the results presented here are also valid for simpler patterns of feedback inhibition, i.e., those that are not "fully-wired." If a pathway with a smaller number of internal feedback interactions is considered, the qualitative results remain the same. To be more specific, the number of sensitivities that are different between pathways with and without overall feedback inhibition may be smaller for pathways with less internal wiring, but the ones that are different remain larger or smaller in the same model as in the fully-wired comparison. This demonstrates the generality of the fully-wired case and the results provide a rationale for the widespread occurrence of overall feedback inhibition in nature.

APPENDIX

One could address the generic questions in this paper because the power-law formalism is systematically structured and is thereby able to represent systems with essentially any type of mechanism, i.e., the representation is mechanism independent. This is in contrast to the Michaelis–Menten formalism, which does not have a well-defined structure [see Savageau (1996)]. One cannot address the generic questions examined in this paper if one insists on using the Michaelis–Menten formalism. The following is an example illustrating why this is the case.

Consider a special case in which one happens to know the specific mechanisms for each reaction in the pathway. For example, assume that all the reactions in common are governed by simple irreversible Michaelis–Menten kinetics, in particular, that the rate law for the degradation of the end product X_n is given by

$$v_n = \frac{V_{mn}X_n}{K_{Mn} + X_n}. \tag{A1}$$

Further assume that the first enzyme has a specific cooperative mechanism with the rate law,

$$v_1 = \frac{V_m X_0^2}{X_0^2 + K_M^2(1 + (X_n^2/K_I^2))}, \tag{A2}$$

and that a mutation-eliminating inhibition by the end product results in the following rate law for the alternative system:

$$v_1 = \frac{V'_m X_0^2}{X_0^2 + K'^2_M}. \tag{A3}$$

In general, the K_M and V_m values will be different in Eqs. A2 and A3, hence primes are used to indicate that the values will be different in the two systems.

If one now generates the conditions for external equivalence, one obtains the following constraint relationships after some differentiation and algebraic manipulation:

$$K'_M = K_M\left\{\frac{1 + (X_{n0}^2/K_I^2)}{1 + 2\dfrac{X_{n0}^2}{K_I^2}\dfrac{K_M^2}{X_0^2}\left(\dfrac{K_{Mn} + X_{n0}}{K_{Mn}}\right)}\right\}^{1/2} \tag{A4}$$

and

$$V'_m = V_m\left\{\frac{X_0^2 + \dfrac{K_M^2(1 + (X_{n0}^2/K_I^2))}{\left[1 + 2\dfrac{X_{n0}^2}{K_I^2}\dfrac{K_M^2}{X_0^2}\left(\dfrac{K_{Mn} + X_{n0}}{K_{Mn}}\right)\right]}}{X_0^2 + K_M^2(1 + X_{n0}^2/K_I^2)}\right\}. \tag{A5}$$

Note that X_{n0} in these expressions has a single positive real solution given by

$$X_{n0} = A + B, \tag{A6}$$

where

$$A = \sqrt[3]{\left[\frac{V_m X_0^2 K_{Mn} K_I^2}{2V_{mn}K_M^2} + \sqrt{\frac{V_m^2 X_0^4 K_{Mn}^2 K_I^4}{4V_{mn}^2 K_M^4} + \frac{K_I^6[V_{mn}K_M^2 + (V_{mn} - V_m)X_0^2]^3}{27V_{mn}^3 K_M^6}}\right]} \tag{A7}$$

and

$$B = \sqrt[3]{\left[\frac{V_m X_0^2 K_{Mn} K_I^2}{2V_{mn}K_M^2} - \sqrt{\frac{V_m^2 X_0^4 K_{Mn}^2 K_I^4}{4V_{mn}^2 K_M^4} + \frac{K_I^6[V_{mn}K_M^2 + (V_{mn} - V_m)X_0^2]^3}{27V_{mn}^3 K_M^6}}\right]}. \tag{A8}$$

If this solution is inserted into the constraint expressions for K'_M and V'_m, one sees that they become even more complex.

These are among the simplest of assumptions regarding the Michaelis–Menten formalism, and one can see how much more complicated this approach is compared to the approach in the power-law formalism [contrast Eqs. A4–A8 with Eqs. 26 and 27 in the text]. The above expressions would be different for different mechanisms, and, when the mechanisms are more complex, the process would become quite impractical. Yet, one obtains the same results for the local behavior.

This work was supported in part by a joint Ph.D. fellowship PRAXIS XXI/BD/9803/96 granted by PRAXIS XXI through Programa Gulbenkian de Doutoramentos em Biologia e Medicina (R.A.), U.S. Public Health Service Grant RO1-GM30054 from the National Institutes of Health (M.A.S.), and U.S. Department of Defense Grant N00014-97-1-0364 from the Office of Naval Research (M.A.S.).

We thank Armindo Salvador for critically reading early versions of this manuscript and making useful comments.

REFERENCES

Alves, R., and M. A. Savageau. 2000a. Extending the method of mathematically controlled comparison to include numerical comparisons. *Bioinformatics.* In press.

Alves, R., and M. A. Savageau. 2000b. Comparing systemic properties of ensembles of biological networks by graphical and statistical methods. *Bioinformatics.* 16:527–533.

Alves, R., and M. A. Savageau. 2000c. Systemic properties of ensembles of metabolic networks: application of graphical and statistical methods to simple unbranched pathways. *Bioinformatics.* 16:534–547.

Costalat, R., and J. Burger. 1996. Effect of enzyme organization on the stability of Yates–Pardee pathways. *Bull. Math. Biol.* 58:719–737.

Demin, O. V., and B. N. Kholodenko. 1993. Structure of metabolic pathways combining the high quality of end product stabilization with high transition rates. *Biochemistry (Moscow).* 58:692–701.

Dibrov, B. F., A. M. Zhabotinsky, and B. N. Kholodenko. 1981. Local stability of the metabolic pathway with end product inhibition. *Biofizica.* 26:590–595.

Dorf, R. C. 1992. Modern Control Systems, 6th ed., Addison-Wesley, Reading, MA.

Frazer, R. A., and W. J. Duncan. 1929. On the criteria for the stability of small motions. *Proc. R. Soc. Lond. A.* 124:642–654.

Goodwin, B. 1963. Temporal Organization in Cells: A Dynamical Theory of Cellular Control Processes, Academic Press, NY.

Hamilton, J. D. 1994. Time Series Analysis, 2nd ed., Princeton University Press, Princeton, NJ.

Hlavacek, W. S., and M. A. Savageau. 1997. Completely uncoupled and perfectly coupled gene expression in repressible pathways. *J. Mol. Biol.* 266:538–558.

Hunding, A. 1974. Limit-cycles in enzyme-systems with nonlinear negative feedback. *Biophys. Struct. Mech.* 1:47–54.

Irvine, D. H. 1991. The method of controlled mathematical comparisons. *In* Canonical Nonlinear Modeling: S-Systems Approach to Understanding Complexity. E. O. Voit, editor. Van Nostrand Reinhold, NY 90–109.

Irvine, D. H., and M. A. Savageau. 1985. Network regulation of the immune response: alternative control points for suppressor modulation of effector lymphocytes. *J. Immunol.* 134:2100–2116.

Koch, A. L. 1967. Metabolic control through reflexive enzyme action. *J. Theor. Biol.* 15:75–102.

Morales, M., and D. McKay. 1967. Biochemical oscillations in "controlled" systems. *Biophys. J.* 7:621–625.

Rapp, P. 1976. Analysis of biochemical phase shift oscillators by a harmonical balancing technique. *J. Math. Biol.* 3:203–224.

Savageau, M. A. 1969. Biochemical systems analysis II: the steady state solution for an *n*-pool system using a power law approximation. *J. Theor. Biol.* 25:370–379.

Savageau, M. A. 1971a. Concepts relating the behavior of biochemical systems to their underlying molecular properties. *Arch. Biochem. Biophys.* 145:612–621.

Savageau, M. A. 1971b. Parameter sensitivity as a criterion for evaluating and comparing the performance of biochemical systems. *Nature.* 229: 542–544.

Savageau, M. A. 1972. The behavior of intact biochemical control systems. *Curr. Top. Cell Reg.* 6:63–130.

Savageau, M. A. 1974. Optimal design of feedback control by inhibition: steady state considerations. *J. Mol. Evol.* 4:139–156.

Savageau, M. A. 1975. Optimal design of feedback control by inhibition: dynamical considerations. *J. Mol. Evol.* 5:199–222.

Savageau, M. A. 1976. Biochemical Systems Analysis: A Study of Function and Design in Molecular Biology. Addison-Wesley, Reading, Mass.

Savageau, M. A. 1996. Power-law formalism: a canonical nonlinear approach to modeling and analysis. *In* World Congress of Nonlinear Analysts 92, Vol. 4 V. Lakshmikantham, editor. Walter de Gruyter Publishers, Berlin. 3323–3334.

Shiraishi, F., and M. A. Savageau. 1992. The tricarboxylic acid cycle in *Dictyostelium discoideum* II. Evaluation of model consistency and robustness. *J. Biol. Chem.* 267:22919–22925.

Thron, C. D. 1991a. The secant condition for instability in biochemical feedback control I: the role of cooperativity and saturability. *Bull. Math. Biol.* 53:383–401.

Thron, C. D. 1991b. The secant condition for instability in biochemical feedback control II: models with upper Hessenberg Jacobian matrices. *Bull. Math. Biol.* 53:403–424.

Umbarger, H. E. 1956. Evidence for a negative-feedback mechanism in the biosynthesis of isoleucine. *Science.* 123:848.

Viniegra-Gonzalez, G. 1973. Stability properties of metabolic pathways with feedback interactions. *In* Biological and Biochemical Oscillators. B. Chance, E. K. Pye, A. K. Ghosh, and B. Hess, editors. Academic Press, New York. 41–59.

Walter, C. F. 1969a. Stability of controlled biological systems. *J. Theor. Biol.* 23:23–38.

Walter, C. F. 1969b. The absolute stability of certain types of controlled biological systems. *J. Theor. Biol.* 23:39–52.

Walter, C. F. 1970. The occurrence and the significance of limit cycle behavior in controlled biological systems. *J. Theor. Biol.* 27:259–272.

Wherry, R. J. 1984. Contributions to Correlational Analysis. Academic Press, Orlando, FL.

Wolfram, S. 1997. Mathematica™: A System for Doing Mathematics by Computer. Addison-Wesley, Menlo Park, CA.

Yates, R. A., and A. B. Pardee. 1956. Control of pyrimidine biosynthesis in *E. coli* by a feedback mechanism. *J. Biol. Chem.* 221:757–770.

Biophysical Journal Volume 80 March 2001 1174–1185

Irreversibility in Unbranched Pathways: Preferred Positions Based on Regulatory Considerations

Rui Alves*†‡ and Michael A. Savageau*

*Department of Microbiology and Immunology, University of Michigan Medical School, 5641 Medical Science Building II, Ann Arbor, Michigan 48109-0620 USA; †Grupo de Bioquimica e Biologia Teoricas, Instituto Rocha Cabral, Calçada Bento da Rocha Cabral 14, 1250 Lisboa, Portugal; and ‡Programa Gulbenkian de Doutoramentos em Biologia e Medicina, Departamento de Ensino, Instituto Gulbenkian de Ciencia, Rua da Quinta Grande 6, 1800 Oeiras, Portugal

ABSTRACT It has been observed experimentally that most unbranched biosynthetic pathways have irreversible reactions near their beginning, many times at the first step. If there were no functional reasons for this fact, then one would expect irreversible reactions to be equally distributed among all positions in such pathways. Since this is not the case, we have attempted to identify functional consequences of having an irreversible reaction early in the pathway. We systematically varied the position of the irreversible reaction in model pathways and compared the resulting systemic behavior according to several criteria for functional effectiveness, using the method of mathematically controlled comparisons. This technique minimizes extraneous differences in systemic behavior and identifies those that are fundamental. Our results show that a pathway with an irreversible reaction located at the first step, and with all other reactions reversible, is on average better than an otherwise equivalent pathway with all reactions reversible, which in turn is on average better than an otherwise equivalent pathway with an irreversible reaction located at any step other than the first. Pathways with an irreversible first reaction and low concentrations of intermediates (one of the primary criteria for functional effectiveness) exhibit the following profile when compared to fully reversible pathways: changes in the concentration of intermediates in response to changes in the level of initial substrate are equally low, the robustness of the intermediate concentrations and of the flux is similar, the margins of stability are similar, flux is more responsive to changes in demand for end product, intermediate concentrations are less responsive to changes in demand for end product, and transient times are shorter. These results provide a functional rationale for the positioning of irreversible reactions at the beginning of unbranched biosynthetic pathways.

INTRODUCTION

Several types of theoretical studies have reported properties of enzymes that could account for their selection during the evolution of metabolic pathways. The simplest type involves determining the distribution of parameter values that produces the maximal catalytic efficiency of an isolated enzyme (Fersht, 1974; Crowley, 1975; Albery and Knowles, 1976; Cornish-Bowden, 1976; Mavrovouniotis et al., 1990; Heinrich and Hoffman, 1991; Peterson, 1992; 1996; Wilhelm et al., 1994; Bish and Mavrovouniotis, 1998; Heinrich and Schuster, 1998). Waley (1964) considered a three-step pathway with reactions described by Michaelis-Menten rate laws and determined the distribution of enzyme concentrations that maximizes flux through the pathway. Similar studies were performed for n-step pathways (Schuster and Heinrich, 1987; Klipp and Heinrich, 1994; Heinrich and Klipp, 1996). Other theoretical studies have dealt with the design of regulatory patterns that, according to multiple criteria, optimize the local behavior of unbranched biosynthetic pathways with n steps and arbitrary mechanisms (Savageau, 1972, 1974, 1975, 1976; Savageau and Jacknow, 1979).

An aspect that has been less thoroughly studied is the distribution of irreversible reactions in unbranched biosynthetic pathways and how this distribution might be related to the optimization of various systemic properties. Although each reaction is in principle reversible, in practice some reactions in a pathway operate far from thermodynamic equilibrium and are effectively irreversible. It has been observed experimentally that, in most cases, unbranched biosynthetic pathways have irreversible reactions near the beginning, many times at the first step, of the pathway (see, e.g., EMP:http://wit.mcs.anl.gov//EMP/).

If there were no functional reasons for irreversible reactions to be at the beginning of a pathway, then one would expect irreversible reactions to be equally distributed among all positions in the pathway. Since this is not the case, we have attempted to identify the functional consequences of having an irreversible reaction early in the pathway. We systematically varied the position of the irreversible reaction in model pathways and compared the resulting systemic behavior according to several criteria for functional effectiveness. The model pathways were represented by a power-law formalism that faithfully captures their nonlinear behavior, independent of mechanistic detail, within a local neighborhood of an arbitrary steady-state operating point. We used the method of mathematically controlled comparison to minimize extraneous differences and to identify fundamental differences. With this approach, we have been able to find a rationale for irreversible reactions at the beginning of unbranched biosynthetic pathways.

Received for publication 19 September 2000 and in final form 19 December 2000.

Address reprint requests to Dr. Michael A. Savageau, University of Michigan, Department of Microbiology and Immunology, 5641 Medical Science II, Ann Arbor, MI 48109-0620. Tel.: 734-764-1466; Fax: 734-763-7163; E-mail: savageau@umich.edu.

© 2001 by the Biophysical Society

0006-3495/01/03/1174/12 $2.00

METHODS

Alternative models and their systemic description

Consider the unbranched biosynthetic pathways depicted in Fig. 1. The initial substrate X_0 is an independent variable with fixed value. The independent variable X_{n+1} represents the cell's demand for the end product X_n. If the cell requires large amounts of X_n, then the value of X_{n+1} will be high; if small amounts of X_n are required, then the value of X_{n+1} will be low. The end product inhibits the first reaction, as has been experimentally observed (Umbarger, 1956; Yates and Pardee, 1956; Monod et al., 1963) and theoretically rationalized (Alves and Savageau, 2000d). The dynamic behavior of such systems can be described by a set of ordinary differential equations.

Assume that the net flux through the pathway is positive (i.e., material is coming into the system from X_0, which is held constant, and exiting the system through X_n). The net positive flux through the reaction immediately before the intermediate X_i (considered the net influx to the pool of X_i) can be accounted for by a single aggregate rate law, representing either the difference between the rate laws for the constituent forward and reverse reactions when the overall reaction is reversible or the rate law for the forward reaction alone when the overall reaction is irreversible. Similarly, the net positive flux through the reaction immediately after the intermediate X_i (considered the net efflux from the pool of X_i) can be represented by a single aggregate rate law.

The dynamical behavior of the models in Fig. 1 can be accurately described in a region about their nominal steady state by using a local S-system representation within the power-law formalism (Savageau, 1969, 1971a, 1976, 1996). For details about different ways to aggregate rate laws and approximate them as S-systems, see Sorribas and Savageau (1989). The resulting equations are the following:

$$\begin{aligned}
\frac{dX_1}{dt} &= \alpha_1 X_0^{g_{10}} X_1^{g_{11}} X_n^{g_{1n}} - \alpha_2 X_1^{g_{21}} X_2^{g_{22}} \\
&\vdots \\
\frac{dX_i}{dt} &= \alpha_i X_{i-1}^{g_{i,i-1}} X_i^{g_{ij}} - \alpha_{i+1} X_i^{g_{i+1,j}} X_{i+1}^{g_{i+1,i+1}} \qquad 0 < i < n \qquad (1) \\
&\vdots \\
\frac{dX_n}{dt} &= \alpha_n X_{n-1}^{g_{n,n-1}} X_n^{g_{nn}} - \alpha_{n+1} X_n^{g_{n+1,n}} X_{n+1}^{g_{n+1,n+1}}
\end{aligned}$$

The aggregate rate law V_i for the influx of X_i is characterized by a multiplicative parameter (rate constant), α_i, which influences the time scale of the reaction and is always positive, and a set of exponential parameters (kinetic orders), g_{ij}, which represents the influence of metabolite X_j on aggregate rate law V_i. If X_j influences the aggregate rate law V_i, either as a reactant or a modulator, and if an increase in the concentration of X_j causes an increase in the rate V_i, then the kinetic order will be positive. If an increase in the concentration of X_j causes a decrease in the rate V_i, then the kinetic order will be negative. If an increase in the concentration of X_j causes neither an increase nor a decrease in the rate V_i, then the kinetic order will be zero. Thus, the positive kinetic orders in Eq. 1 are $g_{i,i-1}$ ($1 \le i \le n+1$), since these are the kinetic orders for substrates of reactions. All other exponents are negative or zero, depending on whether X_i is the product of a reversible ($g_{ii} < 0$) or an irreversible ($g_{ii} = 0$) reaction. The fact that g_{ii} is negative if the reaction is reversible is evident from thermodynamic considerations. If the concentration of the product is increased, the thermodynamic potential across the reversible reaction is reduced and the net flux must decrease. Hence, the kinetic order g_{ii} must be negative to represent this decrease.

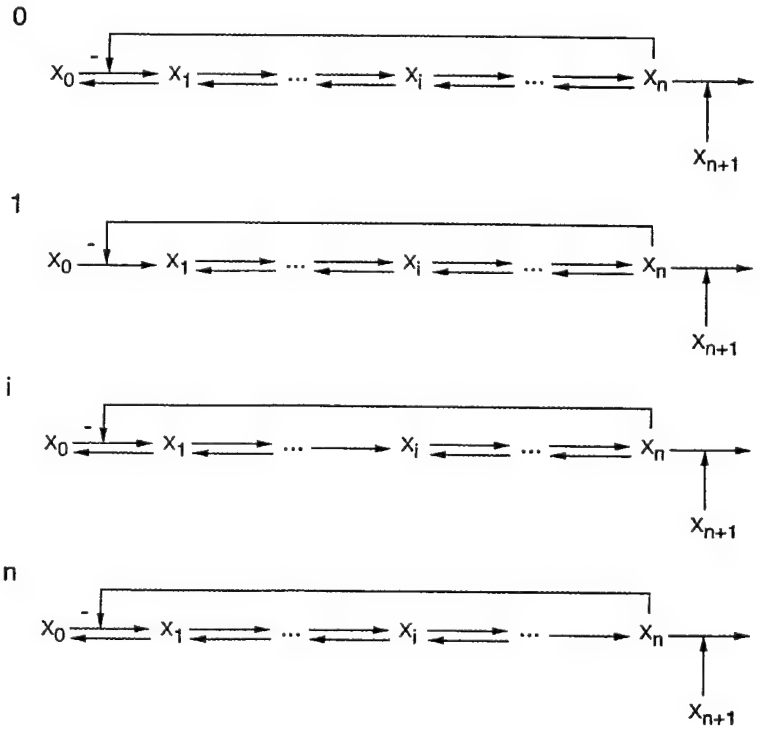


FIGURE 1 Schematic representation of an unbranched biosynthetic pathway subject to control by end-product inhibition. The concentration of the initial substrate X_0 is an independent variable with fixed value; the demand for the end product X_n is represented by X_{n+1}, which also is an independent variable. The reference System 0 has n fully reversible reactions. The alternative systems have one irreversible reaction and the other reactions are identical to the corresponding reactions in the reference system; System 1 has an irreversible reaction at the first position; System i has an irreversible reaction at the ith position; System n has an irreversible reaction at the nth position.

Steady-state solution and key systemic properties

The S-systems describing the dynamic behavior of the models in Fig. 1 can be solved analytically for the steady state (Savageau, 1969, 1971a), where the rates of production and consumption for each metabolite are the same. By equating these rates and taking logarithms of both sides of the resulting equations, one can write the following matrix equation:

$$\begin{bmatrix} b_1 - g_{10}Y_0 \\ b_2 \\ \vdots \\ b_{n-1} \\ b_n + g_{n+1,n+1}Y_{n+1} \end{bmatrix} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} \quad (2)$$

where $Y_i = \log(X_i)$, $b_i = \log(\alpha_{i+1}/\alpha_i)$, and $a_{ij} = g_{ij} - g_{i+1,j}$ for $1 \leq (i, j) \leq n$.

Two types of coefficients, logarithmic gains and parameter sensitivities, can be used to characterize the steady state of such models. Logarithmic gains measure the relative influence of each independent variable on each dependent variable of the model (Savageau, 1971a; Shiraishi and Savageau, 1992). For example,

$$L(X_i, X_0) = \frac{d\log(X_i)}{d\log(X_0)} = \frac{dY_i}{dY_0} \quad (3)$$

measures the percent change in the concentration of intermediate X_i caused by a percentage change in the concentration of the initial substrate X_0. Logarithmic gains provide important information concerning the amplification or attenuation of signals as they are propagated through the system. Parameter sensitivities measure the relative influence of each parameter on each dependent variable of the model (Savageau, 1971b; Shiraishi and Savageau, 1992). For example,

$$S(X_i, p_j) = \frac{d\log(X_i)}{d\log(p_j)} = p_j \frac{dY_i}{dp_j} \quad (4)$$

measures the percentage change in the concentration of intermediate X_i caused by a percentage change in the value of the parameter p_j. Parameter sensitivities provide important information about system robustness, i.e., how sensitive the system is to perturbations in the structural determinants of the system. Because steady-state solutions exist in closed form, we can calculate each of the two types of coefficients simply by taking the appropriate derivatives. Although the mathematical operations involved are the same in each case, it is important to keep in mind that the biological significance of the two types of coefficients is very different.

The local stability of the steady state can be determined by applying the Routh criteria (Dorf, 1992). The magnitude of the two critical Routh conditions can be used to quantify the margin of stability (Savageau, 1976).

Systems should respond quickly to changes in their environment (Savageau, 1975). Thus, another key property of the systems is their temporal response, which was determined as follows. At time zero, each intermediate concentration was set to a value 20% less than its steady-state value. The dynamics were then followed from this initial condition, and the time for all the concentrations to settle to within 1% of their final steady-state value was calculated.

Mathematically controlled comparison

The method of Mathematically Controlled Comparison was specifically developed to make rigorous comparisons of alternative regulatory designs (Savageau, 1972, 1996; Irvine and Savageau, 1985; Alves and Savageau, 2000c, d). This method compares alternative designs for a system that performs a given function and, by using mathematical equivalence constraints to reduce their extraneous differences, determines the irreducible differences between their systemic behaviors. This method requires closed-form solutions for the steady state, which, as noted above, can be obtained with the local S-system representation. Important functional constraints are introduced by equating relevant steady-state properties of the alternative systems being compared. Further analysis (dynamic as well as steady-state) is performed and a profile of ratios is constructed for corresponding results from the alternative systems. In some cases, a ratio can be determined analytically to be less than, equal to, or greater than unity. For example, if the ratio of values for some property P in a reference system to the same property in an alternative system is larger than unity, then the reference system can always be made to have a larger value for P, no matter how large the value for P in the alternative system.

However, if one wishes to know how much greater than unity a given ratio is, then one needs to know actual parameter values. These parameter values are not always available; if they are available, they are not always accurate. Moreover, there are cases in which the ratio can be less than or greater than unity, depending on the specific values for the parameters, so Mathematically Controlled Comparisons that use actual parameter values may lack analytical generality.

In this work we use our method (Alves and Savageau, 2000c), which is a generalization of the original analytical method for making mathematically controlled comparisons; it includes numerical comparisons in which statistical techniques (Alves and Savageau, 2000a) yield results that are general in a statistical sense. We compare the systemic performance of a fully reversible pathway (Fig. 1, System 0) with that of pathways in which only one of the reactions is irreversible (Fig. 1, System 1—System n). We consider all possible positions for the irreversible reaction in pathways with 2 to 7 reactions. The system in which each reaction of the pathway is reversible will be referred to as the reference system or System 0, and the otherwise equivalent system in which the ith reaction of the pathway is irreversible will be referred to as an alternative system or System i. This method also allows direct comparison of System i and System j, each of which has an irreversible reaction but in different positions.

Internal and external equivalence

We are concerned with the irreducible differences in systemic behavior between two pathways of reversible reactions that differ only by the existence of one irreversible reaction in a pathway where the other has a reversible reaction. By irreducible differences we mean differences that persist no matter what the values are for the parameters that define the systems. It is therefore important to ensure that all other changes in systemic behavior are eliminated to the extent possible. To achieve this aim, we shall require that the reference and alternative systems be equivalent from both an internal and external perspective (Savageau, 1972, 1976; Irvine and Savageau, 1985).

By internal equivalence we mean that the values of the corresponding parameters for all the unchanged reactions are the same in both the reference and alternative systems. By external equivalence we mean that systemic behaviors of the reference and alternative systems are made identical, which leads to constraints upon the values for the parameters of the changed reaction. For example, consider the reference system (Fig. 1, System 0) and an alternative system in which the first reaction is irreversible (Fig. 1, System 1). The parameters that characterize the first reaction of the pathway will differ in general between these two systems. The parameters α_1, g_{10}, g_{11}, and g_{1n} of System 0 become the parameters α'_1, g'_{10}, $g'_{11} = 0$, and g'_{1n} of System 1. Since we wish to determine the necessary systemic effects that are due to the change from reversibility to irreversibility, we shall specify values for the parameters α'_1, g'_{10}, and g'_{1n} that eliminate as many extraneous systemic effects as possible. This is accomplished by deriving mathematical expressions for a given steady-state property in each of the two models, equating these expressions to produce a constraint equation, and then solving the constraint equation for one of the primed parameters in terms of the unprimed parameters. When all

primed parameters have been specified in this fashion, there will be no more degrees of freedom with which to make systemic properties equivalent between the two models, and the two systems will be maximally equivalent from an external perspective.

Calculating the constraints for external equivalence

We require the reference and alternative systems in Fig. 1 to have the same steady-state logarithmic gains with respect to the initial substrate of the pathway and the same concentrations (and thus flux). These two types of constraints are sufficient to fix the two primed parameters of the irreversible reaction when its position is beyond the first step.

When the position of the irreversible reaction is at the first step, there are three primed parameters that need to be fixed (see previous section). For the third constraint we require the reference and alternative systems in Fig. 1 to have the same sensitivity of the concentrations with respect to changes in the parameter α_1. This constraint is preferred over other possibilities because the reference system and alternative system will then exhibit the smallest number of systemic differences, which is the objective in a mathematically controlled comparison. One could choose a different systemic property to form the third constraint. However, the reference system and alternative system would then exhibit a larger number of systemic differences, some of which could be eliminated by the choice of the preferred constraint.

Thus, the following system of algebraic equations is solved to obtain the analytic constraints for the primed parameters of the irreversible reaction at the ith step:

$$L(X_i, X_0)_{\text{Reference}} = L(X_i, X_0)_{\text{Alternative}} \quad 1 \le i \le n \tag{5a}$$

$$S(X_i, \alpha_1)_{\text{Reference}} = S(X_i, \alpha_1)_{\text{Alternative}} \quad i = 1 \tag{5b}$$

$$\log[X_i]_{\text{Reference}} = \log[X_i]_{\text{Alternative}} \quad 1 \le i \le n \tag{5c}$$

By constraining one of the logarithmic gains (Eq. 5a), all of them are constrained. This allows us to fix the kinetic order $g'_{i,i-1}$. When the irreversible reaction occurs at the first step, the additional constraint (Eq. 5b) allows us to fix the kinetic order g'_{1n}. By constraining one of the concentrations (Eq. 5c), all of them, as well as the steady-state flux, are constrained. This allows us to fix the rate constant α'_i.

The parametric constraints obtained by solving Eq. 5 have the following form:

$$\begin{aligned} g'_{i,i-1} &= g_{i,i-1} f_i(g, n) \\ g'_{1n} &= g_{1n} + f_n(g, n) \\ \log(\alpha'_i) &= f_\alpha(\alpha, g, n) \end{aligned} \tag{6}$$

where the parameters α and g in the functions f are intended to represent a set of rate constants and kinetic orders that depend both on the length of the pathway and on the systems being considered. The specific forms of these constraints are presented in the Appendix for $n = 2$ to $n = 7$.

Numerical analysis

The analytical results give qualitative information that characterizes the effect of irreversibility in the systems of Fig. 1. To obtain quantitative information, one must introduce specific values for the parameters and compare systems. For this purpose we have randomly generated a large ensemble of parameter sets and selected 5000 of these sets that define systems consistent with various physical and biochemical constraints. These constraints include mass balance, low concentrations of intermediates and small changes in their values to minimize utilization of the solvent capacity in the cell, small values for parameter sensitivities so as to desensitize the system to spurious fluctuations affecting its structure, and stability margins large enough to ensure local stability of the systems. A detailed description of these methods can be found in Alves and Savageau (2000b). Mathematica (Wolfram, 1997) was used for all the numerical procedures.

Density of ratios plot

To interpret the ratios that result from our analysis, we use Density of Ratios plots as defined in Alves and Savageau (2000a). The primary density plots from the raw data have the magnitude for some property of the reference system on the x-axis and the corresponding ratio of magnitudes (reference system to alternative system) on the y-axis. The primary plot can be viewed as a list of 5000 paired values that can be ordered with respect to the reference magnitude to form a list L_1 in which the first pair has the lowest measured value for property P in the reference model, the second has the second lowest, and so on. Secondary density plots are constructed from the primary plots by the use of moving quantile techniques with a window size of 500. The procedure is as follows. One collects the first 500 ratios from the list L_1, calculates the quantile of interest for this sample, and pairs this number ⟨R⟩ with the median value of the corresponding P values of the reference model, denoted ⟨P⟩. One advances the window by one position, collects ratios 2 through 501, calculates ⟨R⟩, and pairs it with the corresponding ⟨P⟩ value and continues in this manner until the last ratio from the list L_1 is used for the first time. This procedure generates a second list L_2 and the corresponding secondary plot. The slope in the secondary plot measures the degree of correlation between the quantities plotted on the x- and y-axes.

Mathematically controlled comparison

Several criteria are considered to determine the functional effectiveness of unbranched biosynthetic pathways (Savageau, 1976; Alves and Savageau, 2000d). The systems being compared will be equal on the bases of the first two criteria because of external equivalence constraints, whereas they will differ with respect to the remaining five criteria.

1. The concentration of intermediates should be low, because otherwise it would tax the limited solvent capacity of the cell and potentially interfere in a nonspecific way with unrelated reactions (e.g., Atkinson, 1969; Savageau, 1972; Srere, 1987; see Levine and Ginsburg, 1985, for a general discussion of the subject from different perspectives). Due to the conditions for external equivalence that we shall impose, the concentrations of the corresponding intermediates will be the same for all comparable systems being examined.
2. The changes in concentration of intermediates caused by changes in the initial substrate should be small. This also will ensure that the solvent capacity is not exceeded when the concentration of intermediates changes. Again, due to the conditions for external equivalence, the corresponding logarithmic gains will be the same for all the systems being examined. These changes are quantified by means of the logarithmic-gain factors $L(X_i, X_0)$ as defined in Eq. 3.
3. The systems should be robust, i.e., the concentrations and flux should be insensitive to changes in the parameters that define the structure of the system (Savageau, 1971b; Shiraishi and Savageau, 1992). If these sensitivities are high, then small fluctuations in parameter values (e.g., due to physical changes such as temperature or to errors in replication, transcription, or translation) would lead to large deviations from the normal behavior of the system. These changes are quantified by means of the parameter sensitivities $S(X_i, p_j)$ and $S(V, p_j)$ as defined in Eq. 4. Aggregate sensitivities for intermediate concentrations and flux are defined as follows: $S(X_i) = \sqrt{\Sigma_j S(X_i\ p_j)^2}$ and $S(V) = \sqrt{\Sigma_j S(V, p_j)^2}$.
4. The systems should have a steady state that is dynamically stable following small perturbations in the concentration variables, otherwise they would be dysfunctional, i.e., unable to maintain homeostasis in the

face of spurious perturbations. Furthermore, the margins of stability should be sufficiently large that changes in parameter values will not produce an unstable steady state. There are n Routh conditions that determine whether the steady state of a system with n variables will be stable. The margins of stability are quantified by the size of the critical Routh conditions, which are the last two (Savageau, 1976; Hlavacek and Savageau, 1997).

5. The flux through the pathway should be highly responsive to changes in the demand for end product. This ensures that the amount of material flowing through the pathway is tightly coupled to the needs of cellular metabolism. This criterion is quantified by the logarithmic gain $L(V, X_{n+1})$, as defined in Eq. 3.
6. The changes in concentration of the intermediates caused by changes in demand for the end product should be small. This ensures that the depletion of end product is minimized when there is an increase in demand. It also ensures that the solvent capacity is not exceeded by the intermediates when demand for the end product changes. These changes are quantified by means of the logarithmic-gain factors $L(X_n, X_{n+1})$ and $L(X_i, X_{n+1})$ as defined in Eq. 3.
7. The systems should respond quickly to changes in their environment, i.e., they should have short transient times (Savageau, 1975). Organisms harboring systems with a sluggish response to change will be at a disadvantage when competing with other organisms in a rapidly changing environment. Transient time will be measured as the time it takes the system to return to its steady state after a small perturbation in concentrations.

RESULTS

In all the results described below, the reference and alternative systems have the same steady-state values for the flux through the pathway, the same concentrations of the corresponding metabolites, and the same logarithmic gains for pathway flux and for metabolite concentrations in response to changes in the initial substrate. These equivalent behaviors are a direct consequence of the constraints for internal and external equivalence, as described above in Methods. The reference and alternative systems differ on the basis of their robustness, margin of stability, response to demand for end product, and transient time.

Robustness

We compare the robustness of the reference system having all reversible reactions with that of an otherwise equivalent alternative system having one irreversible reaction in all possible positions. In most cases, symbolic analysis is sufficient to determine whether the ratio of a given parameter sensitivity in the reference system to the corresponding sensitivity in the alternative system is larger or smaller than 1; in the remaining cases, symbolic analysis is incapable of determining the value for the ratio because it depends on the specific values of the parameters. Results of the symbolic analysis are summarized in Table 1 for pathways of length 2 to 7. The following patterns can be observed in the data.

The reference system is always more robust than the alternative system with an irreversible synthesis of the end product, because the ratios of parameter sensitivities are all less than or equal to 1. As the position of the irreversible reaction approaches the beginning of the pathway, the number of sensitivities that are equal in the systems being compared decreases. The concentration of the product of the irreversible reaction is always more sensitive to parameter changes than the product of the corresponding reversible reaction in the reference system.

In general, numerical methods are needed to decide which systems are more robust because this cannot be done by examining just the symbolic sensitivities. The numerical results in Fig. 2 *A* show that the aggregate sensitivity of X_i to parameters is on average the same in the reference and alternative systems if X_i is surrounded by reversible reactions. If either the reaction that produces or the reaction that consumes X_i is irreversible, then that concentration is on average more robust in the reference system. Fig. 2 *B* shows that, on average, the reference System 0 has smaller aggregate sensitivities for flux than alternative Systems i. However, these differences are only significant for alternative Systems 1 and n.

Margin of stability

Comparing System 0 with System i shows that the stability margins for systems with 2 reactions are always larger in a reference System 0. For systems with 3 to 7 reactions, these margins can be larger in either system. Direct comparison of System i with System j shows that the stability margins can be larger in either system, depending on the parameter values.

Numerical results show that, on average, the reference System 0 has larger margins of stability than alternative Systems i ($i > 1$). Numerical results also show that, on average, the reference System 0 has larger margins of stability than System 1, although the differences are insignificant (Fig. 2 *C*).

Response to demand for end product

Symbolic comparisons with the reference system show that the flux through System 1 is more responsive to changes in the demand for end product than is the flux through System 0. However, for $i > 1$, the flux through System 0 is more responsive to changes in the demand for end product than is the flux through System i. This demonstrates that, with respect to this systemic property, System 1 is better than System 0 and better than any of the other alternatives. Direct comparison of Systems i and j with respect to this systemic property reveals additional information. If $i, j > 1$, then the flux through Systems i and j is equally responsive to changes in the demand for end product.

Numerical results (Fig. 2 *D*) show that average differences between the reference System 0 and alternative System 1 are about 120%, whereas the differences between the

TABLE 1 Comparison of parameter sensitivities for the reference and alternative systems as a function of pathway length and of position for the irreversible step in the pathway

	$n=2$				$n=3$				$n=4$				$n=5$				$n=6$				$n=7$			
	>1	<1	=1	?	>1	<1	=1	?	>1	<1	=1	?	>1	<1	=1	?	>1	<1	=1	?	>1	<1	=1	?
1st reaction irreversible ($i=1$)																								
V	3	3	2	0	6	3	2	0	6	6	2	0	12	0	2	3	1	17	2	0	0	21	2	0
X_1	0	4	2	2	0	7	2	2	0	6	2	5	1	12	2	2	16	0	2	2	19	0	2	2
X_2	5	1	2	0	3	4	2	2	3	2	2	7	4	11	2	0	12	4	2	2	15	4	2	2
X_3	—	—	—	—	8	1	2	0	11	1	2	0	8	7	2	0	9	7	2	2	12	7	2	2
X_4	—	—	—	—	—	—	—	—	8	4	2	0	11	2	2	2	7	9	2	2	10	9	2	2
X_5	—	—	—	—	—	—	—	—	—	—	—	—	13	2	2	0	3	13	2	2	6	13	2	2
X_6	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	17	2	0	4	15	2	2
X_7	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0	21	2	0
2nd reaction irreversible ($i=2$)																								
V	2	1	5	0	5	1	5	0	8	1	5	0	11	1	5	0	3	12	5	0	3	15	5	0
X_1	0	3	5	0	3	3	5	0	3	6	5	0	9	3	5	0	3	12	5	0	3	15	5	0
X_2	0	3	5	0	0	4	5	2	0	2	5	7	0	10	5	2	13	0	5	2	16	0	5	2
X_3	—	—	—	—	3	3	5	0	3	6	5	0	3	7	5	2	10	3	5	2	13	3	5	2
X_4	—	—	—	—	—	—	—	—	6	3	5	0	6	4	5	2	7	6	5	2	10	6	5	2
X_5	—	—	—	—	—	—	—	—	—	—	—	—	9	3	5	0	3	10	5	2	6	10	5	2
X_6	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	3	12	5	0	3	13	5	2
X_7	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	3	15	5	0
3rd reaction irreversible ($i=3$)																								
V	—	—	—	—	2	1	8	0	5	1	8	0	8	1	8	0	3	9	8	0	2	13	8	0
X_1	—	—	—	—	0	3	8	0	0	6	8	0	6	3	8	0	3	9	8	0	3	12	8	0
X_2	—	—	—	—	0	3	8	0	0	6	8	0	6	3	8	0	3	9	8	0	3	12	8	0
X_3	—	—	—	—	0	3	8	0	0	6	8	0	0	7	8	2	10	0	8	2	13	0	8	2
X_4	—	—	—	—	—	—	—	—	3	3	8	0	3	4	8	2	7	3	8	2	10	3	8	2
X_5	—	—	—	—	—	—	—	—	—	—	—	—	6	3	8	0	4	6	8	2	7	6	8	2
X_6	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	3	9	8	0	3	10	8	2
X_7	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2	13	8	0
4th reaction irreversible ($i=4$)																								
V	—	—	—	—	—	—	—	—	2	1	11	0	5	1	11	0	7	2	11	0	2	10	11	0
X_1	—	—	—	—	—	—	—	—	0	3	11	0	4	3	11	0	5	4	11	0	7	5	11	0
X_2	—	—	—	—	—	—	—	—	0	3	11	0	4	3	11	0	4	5	11	0	6	6	11	0
X_3	—	—	—	—	—	—	—	—	0	3	11	0	4	3	11	0	6	3	11	0	9	3	11	0
X_4	—	—	—	—	—	—	—	—	0	3	11	0	0	4	11	2	7	0	11	2	10	0	11	2
X_5	—	—	—	—	—	—	—	—	—	—	—	—	4	3	11	0	4	3	11	2	7	3	11	2
X_6	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2	7	11	0	3	7	11	2
X_7	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2	10	11	0
5th reaction irreversible ($i=5$)																								
V	—	—	—	—	—	—	—	—	—	—	—	—	2	1	14	0	2	4	14	0	2	7	14	0
X_1	—	—	—	—	—	—	—	—	—	—	—	—	0	3	14	0	2	4	14	0	6	1	14	2
X_2	—	—	—	—	—	—	—	—	—	—	—	—	0	3	14	0	1	3	14	1	5	2	14	2
X_3	—	—	—	—	—	—	—	—	—	—	—	—	0	3	14	0	1	3	14	2	5	2	14	2
X_4	—	—	—	—	—	—	—	—	—	—	—	—	0	3	14	0	2	2	14	2	5	2	14	2
X_5	—	—	—	—	—	—	—	—	—	—	—	—	0	3	14	0	2	2	14	2	5	2	14	2
X_6	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2	4	14	0	0	7	14	2
X_7	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2	10	14	0
6th reaction irreversible ($i=6$)																								
V	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2	1	17	0	5	1	17	0
X_1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0	3	17	0	2	2	17	2
X_2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0	3	17	0	2	2	17	2
X_3	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0	3	17	0	3	1	17	2
X_4	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0	3	17	0	2	2	17	2
X_5	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0	3	17	0	4	0	17	2
X_6	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0	3	17	0	4	0	17	2
X_7	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	5	1	17	

TABLE 1 Continued

	$n = 2$				$n = 3$				$n = 4$				$n = 5$				$n = 6$				$n = 7$			
	>1	<1	=1	?	>1	<1	=1	?	>1	<1	=1	?	>1	<1	=1	?	>1	<1	=1	?	>1	<1	=1	?
7th reaction irreversible ($i = 7$)																								
V	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2	1	20	0
X_1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0	3	20	0
X_2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0	3	20	0
X_3	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0	3	20	0
X_4	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0	3	20	0
X_5	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0	3	20	0
X_6	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0	3	20	0
X_7	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0	3	20	0

The sensitivities of the steady-state flux (V) through the pathway and of the steady-state concentrations (X_i) are calculated with respect to each of the parameters in both the reference system and the alternative system. The ratio of a given sensitivity in the reference system relative to the corresponding sensitivity in the alternative system is determined to be greater than one, less than one, equal to one, or indeterminate. The number of reactions in the pathway, n, varies from 2 to 7. The position of the irreversible reaction in the pathway, i, varies from 1 to n. The ratios are the values of the parameter sensitivities for reference System 0 relative to those for alternative Systems i (see Fig. 1). Column legend: >1, number of sensitivities that are larger in reference System 0; <1, number of sensitivities that are smaller in reference System 0; =1, number of sensitivities that are the same in both systems under comparison; ?, number of sensitivities that can be larger in either system, depending on parameter values. For example, the number 5 at the 3rd row, 1st column position of the $i = 1$, $n = 2$ section of the table means that there are five different parameters in a two-step pathway for which the sensitivities of X_2 are larger in System 0 than in System 1.

reference System 0 and alternative Systems i ($i > 1$) are, on average, less than 2%.

The end-product concentration in System 1 is less responsive to changes in the demand for end product than is the end product in System 0. However, for $i > 1$, the end product concentration in System 0 is less responsive to changes in the demand for end product than is the end product in System i. Again, System 1 is better than System 0 and better than any of the other alternatives.

Numerical results (Fig. 2 *E*) show that average differences between the reference System 0 and alternative System 1 can be between 50 and 100%, whereas the differences between the reference System 0 and alternative Systems i ($i > 1$) are, on average, much smaller (2–8%).

Transient time

There is no explicit solution for the dynamic equations given in Eq. 1 that would allow one to determine symbolically the transient responses of the various systems in Fig. 1. The numerical results in Fig. 2 *F* show that the transient time for alternative Systems i ($i < n$) is, on average, smaller than that for the reference System 0, whereas the transient time for alternative System n is larger than that for the reference System 0. A direct comparison of System i and System j ($i, j \neq n$) shows that the transient time can be larger in either system, depending on the length of the pathway (data not shown).

Correlations between ratios and systemic properties

The aggregate sensitivities of the concentrations in System i on average approach those in System 0 as the concentrations of intermediates decrease, i.e., the ratio of aggregate sensitivities approaches 1 (Fig. 2 *A*). The ratio for aggregate sensitivities of flux in System 0 and System 1 also approaches 1, whereas the same ratio in System 0 and Systems i ($i > 1$) decreases away from 1 (Fig. 2 *B*). Thus, the differences in robustness (criterion 3) in System 0 and System 1 become less significant, whereas the differences in System 0 and Systems i ($i > 1$) become more significant at low concentrations of intermediates, which is our first criterion for functional effectiveness.

The ratios involving the critical margins of stability can be positively or negatively correlated with the concentrations of intermediates, depending on the particular comparison (Fig. 2 *C*). There is no general pattern apparent in this panel, so these correlations provide no further information regarding criterion 4.

The ratios for System 0 relative to System 1 of logarithmic gains in flux with respect to changes in the demand for end product are positively correlated with low concentrations of intermediates, although the slope for this correlation is small. The same ratios, but for System 0 relative to System i ($i > 1$), are negatively correlated with low concentrations of intermediates, although the slope for this correlation is also small (Fig. 2 *D*). Thus, the differences in responsiveness of flux to changes in demand for end product (criterion 5) in System 0 and System 1, and in System 0 and System i ($i > 1$), become more significant at low concentrations of intermediates.

The ratios involving logarithmic gains in end product concentration with respect to changes in the demand for end product are positively correlated with the concentrations of intermediates (Fig. 2 *E*). Thus, the differences in depletion of end product following an increase in demand for end product (criterion 6) in System 0 and System 1 become less significant at low concentrations of intermediates.

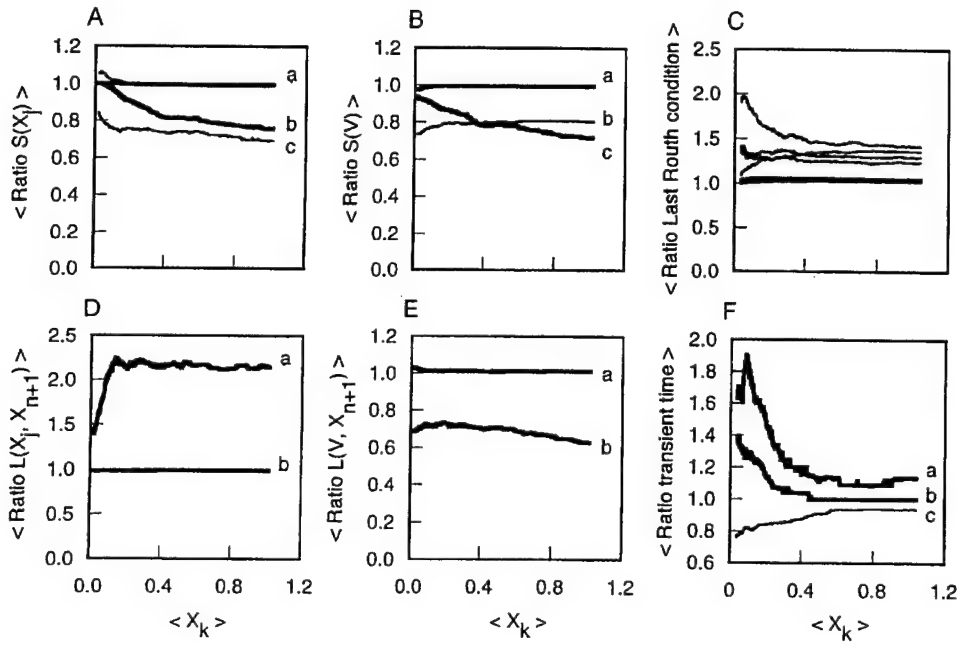


FIGURE 2 Typical correlation curves between ratios of magnitudes in reference System 0 relative to those in alternative Systems i versus concentrations of intermediates. The data, which are generated by changing all of the parameter values randomly within the constraints described in the Methods section, are displayed in a density of ratios plot (Alves and Savageau, 2000a). The y-axis indicates which of two systems on average has the larger magnitude; the x-axis indicates how this difference changes as a function of the concentration of intermediates (see criterion 1 in the text). The subscripts j and k refer to arbitrary pathway intermediates, which have different concentrations in general. We have made individual plots for each pathway length and combination of intermediates. However, since the trends observed for different pathway lengths and intermediates are the same, we show only representative examples. (*A*) Ratios of aggregate sensitivities of concentrations: *a*, aggregate sensitivities of metabolites that have both their production and consumption catalyzed by reversible reactions; *b* and *c*, aggregate sensitivities of metabolites that have either their production or consumption catalyzed by an irreversible reaction. (*B*) Ratios of aggregate sensitivities of flux: *a*, ratio for reference System 0 relative to alternative Systems i ($1 < i < n$); *b*, ratio for reference System 0 relative to alternative System n; *c*, ratio for reference System 0 relative to alternative System 1. (*C*) Ratios of critical criteria for local stability. (*D*) Ratios of logarithmic gains in concentration with respect to changes in demand for the end product: *a*, ratio for reference System 0 relative to alternative System 1; *b*, ratio for reference System 0 relative to alternative System i ($i > 1$). (*E*) Ratios of logarithmic gains in flux with respect to changes in demand for the end product: *a*, ratio for reference System 0 relative to alternative Systems i ($i > 1$); *b*, ratio for reference System 0 relative to alternative System 1. (*F*) Ratios of transient times: *a* and *b*, ratio for reference System 0 relative to two different alternative Systems i ($i < n$); *c*, ratio for reference System 0 relative to alternative System n.

The ratios involving transient times are inversely correlated with the concentrations of intermediates (Fig. 2 *F*). Thus, the difference in transient times (criterion 7) in System 0 and Systems i ($i < n$) increases as the concentration of intermediates decreases, whereas this difference in System 0 and System n decreases.

DISCUSSION

We analyzed the effect of having an irreversible reaction at different positions in an unbranched biosynthetic pathway with all other reactions being reversible. We also analyzed the effect of having a reversible reaction at different positions in pathways with all other reactions being irreversible (data not shown). The results are qualitatively similar; namely, the best position for the single irreversible reaction is at the beginning of the pathway, whereas the best position for the single reversible reaction is at the end of the pathway. The method used for our analysis, mathematically controlled comparisons, often allows one to obtain symbolic (and thus general) results when comparing systemic properties of alternative models. When this is not possible, the method also can be used numerically to obtain results that are general in a statistical sense. Comparisons were made based on functional effectiveness, as judged by the seven quantitative criteria described in detail in the Methods section.

In this work we have found a limited number of symbolic comparisons whose conclusions do not depend on the specific values of the parameters. The reference pathway with all reactions fully reversible (System 0) is more robust to perturbations in the values of the parameters (criterion 3) than is an otherwise equivalent alternative pathway with an irreversible synthesis of end product. Also, when comparing reference System 0 with alternative Systems i ($i \neq 1$), where reaction i is irreversible, the flux through System 0 is more responsive to changes in the demand for end product (cri-

terion 5), whereas the concentrations of its intermediates are less responsive (criterion 6). On the other hand, the flux is more responsive (criterion 5) and the concentrations are less responsive (criterion 6) to the demand for end product in System 1 than in System 0. Taken together, these results imply that reference System 0 is superior to alternative System n on the bases of criteria 3, 5, and 6, superior to alternative Systems i ($i \neq 1, n$) on the bases of criteria 5 and 6, but inferior to alternative System 1 on the bases of criteria 5 and 6. Not much can be said analytically about the comparison of these systems based on other criteria.

Additional conclusions that are general in a statistical sense can be obtained by means of numerical comparisons. These indicate that the reference System 0 is, on average, better than or similar to the alternative Systems i ($i > 1$) on the bases of all the criteria except transient time (criterion 7). These numerical comparisons also indicate that the alternative System 1 is, on average, better than or similar to the reference System 0 on the bases of all the criteria except some components of robustness (criterion 3). The differences in value for those components that favor reference System 0 over alternative System 1 are less significant when the systems are optimized according to criterion 1 than when these systems are not so optimized. Thus, alternative System 1 is, on average, better than or similar to all other systems under the following conditions: The concentrations of intermediates are equally low (criterion 1). The logarithmic gains in concentration with respect to change in the level of initial substrate also are equally low (criterion 2). The robustness of all the intermediates, with one exception, is similar. Although, as noted above, the first intermediate and the flux are less robust in System 1, these differences are less significant when criterion 1 is satisfied (criterion 3). The margins of stability are similar (criterion 4). Flux is more responsive to changes in demand for end product (criterion 5). Concentrations of intermediates are less responsive to changes in demand for end product (criterion 6). Transient times are shorter (criterion 7).

The combination of analytical and numerical results presented in this paper provides a functional rationale for why irreversible reactions are found predominantly at the beginnings of unbranched biosynthetic pathways.

APPENDIX

Parametric constraints for external equivalence. The number of reactions in the pathway is n, where n varies from 2 to 7. The position of the irreversible reaction in the pathway is i, where i varies from 1 to n. An extra constraint, $g'_{10} = g_{10}$, is common to all cases when the irreversible reaction is in the first position, i.e., when i = 1.

$n = 2$

$$i = 1: \log[\alpha'_1] = \log[\alpha_1] - g_{11}\log[\alpha_2/\alpha_3]/g_{21}; \quad g'_{12} = g_{12} + g_{11}(g_{32} - g_{22})/g_{21}$$

$$i = 2: \log[\alpha'_2] = (g_{32}\log[\alpha_2] - g_{22}\log[\alpha_3])/(g_{32} - g_{22}); \quad g'_{21} = g_{21}g_{32}/(g_{32} - g_{22})$$

$n = 3$

$$i = 1: \log[\alpha'_1] = \log[\alpha_1] - g_{11}(g_{32}\log[\alpha_2/\alpha_4] - g_{22}\log[\alpha_3/\alpha_4])/(g_{21}g_{32});$$

$$g'_{13} = g_{13} + g_{11}(g_{32}g_{43} - g_{22}(g_{43} - g_{33}))/(g_{21}g_{32})$$

$$i = 2: \log[\alpha'_2] = (g_{43}(g_{32}\log[\alpha_2] - g_{22}\log[\alpha_3]) + g_{22}g_{33}\log[\alpha_4])/(g_{32}g_{43} - g_{22}(g_{43} - g_{33}));$$

$$g'_{21} = g_{21}g_{32}g_{43}/(g_{32}g_{43} - g_{22}(g_{43} - g_{33}))$$

$$i = 3: \log[\alpha'_3] = (g_{43}\log[\alpha_3] - g_{33}\log[\alpha_4])/(g_{43} - g_{33}); \quad g'_{32} = g_{32}g_{43}/(g_{43} - g_{33})$$

$n = 4$

$$i = 1: \log[\alpha'_1] = \log[\alpha_1] - g_{11}\frac{g_{32}g_{43}\log[\alpha_2/\alpha_5] - g_{22}(g_{43}\log[\alpha_3/\alpha_5] - g_{33}\log[\alpha_4/\alpha_5])}{g_{21}g_{32}g_{43}};$$

$$g'_{14} = g_{14} + g_{11}(g_{22}g_{33}(g_{54} - g_{44}) + g_{43}g_{54}(g_{32} - g_{22}))/(g_{21}g_{32}g_{43})$$

$$i = 2: \log[\alpha'_2] = \frac{g_{43}g_{54}(g_{32}\log[\alpha_2] - g_{22}\log[\alpha_3]) + g_{22}g_{33}(g_{54}\log[\alpha_4] - g_{44}\log[\alpha_5])}{g_{32}g_{43}g_{54} - g_{22}(g_{43}g_{54} - g_{33}(g_{54} - g_{44}))};$$

$$g'_{21} = g_{21}g_{32}g_{43}g_{54}/(g_{32}g_{43}g_{54} - g_{22}(g_{43}g_{54} - g_{33}(g_{54} - g_{44})))$$

$$i = 3\text{: } \log[\alpha'_3] = (g_{54}(g_{43}\log[\alpha_3] - g_{33}\log[\alpha_4]) + g_{33}g_{44}\log[\alpha_5])/(g_{43}g_{54} - g_{33}(g_{54} - g_{44}));$$

$$g'_{32} = g_{32}g_{43}g_{54}/(g_{43}g_{54} - g_{33}(g_{54} - g_{44}))$$

$$i = 4\text{: } \log[\alpha'_4] = (g_{54}\log[\alpha_4] - g_{44}\log[\alpha_5])/(g_{54} - g_{44}); \quad g'_{43} = g_{43}g_{54}/(g_{54} - g_{44})$$

$n = 5$

$$i = 1\text{: } \log[\alpha'_1] = \log[\alpha_1] - g_{11}\frac{(g_{32}g_{43}g_{54}\log[\alpha_2/\alpha_6] - g_{22}g_{43}g_{54}\log[\alpha_3/\alpha_6] + g_{22}g_{33}g_{54}\log[\alpha_4/\alpha_6] - g_{22}g_{33}g_{44}\log[\alpha_5/\alpha_6])}{g_{21}g_{32}g_{43}g_{54}};$$

$$g'_{15} = g_{15} + g_{11}\frac{g_{22}g_{33}(g_{54}g_{65} - g_{44}(g_{65} - g_{55})) + g_{43}g_{54}g_{65}(g_{32} - g_{22})}{g_{21}g_{32}g_{43}g_{54}}$$

$$i = 2\text{: } \log[\alpha'_2] = \frac{(g_{32}g_{43}g_{54}g_{65}\log[\alpha_2] - g_{22}g_{43}g_{54}g_{65}\log[\alpha_3] + g_{22}g_{33}g_{54}g_{65}\log[\alpha_4] - g_{22}g_{33}g_{44}g_{65}\log[\alpha_5] + g_{22}g_{33}g_{44}g_{55}\log[\alpha_6])}{g_{32}g_{43}g_{54}g_{65} - g_{22}(g_{43}g_{54}g_{65} - g_{33}(g_{54}g_{65} - g_{44}(g_{65} - g_{55})))};$$

$$g'_{21} = g_{21}g_{32}g_{43}g_{54}g_{65}/(g_{32}g_{43}g_{54}g_{65} - g_{22}(g_{43}g_{54}g_{65} - g_{33}(g_{54}g_{65} - g_{44}(g_{65} - g_{55}))))$$

$$i = 3\text{: } \log[\alpha'_3] = \frac{g_{43}g_{54}g_{65}\log[\alpha_3] - g_{33}g_{54}g_{65}\log[\alpha_4] + g_{33}g_{44}(g_{65}\log[\alpha_5] - g_{55}\log[\alpha_6])}{g_{43}g_{54}g_{65} - g_{33}(g_{54}g_{65} - g_{44}(g_{65} - g_{55}))};$$

$$g'_{32} = g_{32}g_{43}g_{54}g_{65}/(g_{43}g_{54}g_{65} - g_{33}(g_{54}g_{65} - g_{44}(g_{65} - g_{55})))$$

$$i = 4\text{: } \log[\alpha'_4] = (g_{54}g_{65}\log[\alpha_4] - g_{44}(g_{65}\log[\alpha_5] - g_{55}\log[\alpha_6]))/(g_{54}g_{65} - g_{44}(g_{65} - g_{55}));$$

$$g'_{43} = g_{43}g_{54}g_{65}/(g_{54}g_{65} - g_{44}(g_{65} - g_{55}))$$

$$i = 5\text{: } \log[\alpha'_5] = (g_{65}\log[\alpha_5] - g_{55}\log[\alpha_6])/(g_{65} - g_{55}); \quad g'_{54} = g_{54}g_{65}/(g_{65} - g_{55})$$

$n = 6$

$$i = 1\text{: } \log[\alpha'_1] = \log[\alpha_1] - g_{11}\left(\frac{\begin{pmatrix} g_{32}g_{43}g_{54}g_{65}\log[\alpha_2/\alpha_7] - g_{22}g_{43}g_{54}g_{65}\log[\alpha_3/\alpha_7] \\ + g_{22}g_{33}g_{54}g_{65}\log[\alpha_4/\alpha_7] - g_{22}g_{33}g_{44}g_{65}\log[\alpha_5/\alpha_7] \\ + g_{22}g_{33}g_{44}g_{55}\log[\alpha_6/\alpha_7] \end{pmatrix}}{g_{21}g_{32}g_{43}g_{54}g_{65}}\right);$$

$$g'_{16} = g_{16} + g_{11}\frac{g_{32}g_{43}g_{54}g_{65} - g_{22}(g_{43}g_{54}g_{65} - g_{33}(g_{54}g_{65} - g_{44}(g_{65}g_{76} - g_{55}(g_{76} - g_{66}))))}{g_{21}g_{32}g_{43}g_{54}g_{65}}$$

$$i = 2\text{: } \log[\alpha'_2] = \frac{\begin{pmatrix} g_{32}g_{43}g_{54}g_{65}g_{76}\log[\alpha_2] - g_{22}g_{43}g_{54}g_{65}g_{76}\log[\alpha_3] + g_{22}g_{33}g_{54}g_{65}g_{76}\log[\alpha_4] \\ - g_{22}g_{33}g_{44}g_{65}g_{76}\log[\alpha_5] + g_{22}g_{33}g_{44}g_{55}g_{76}\log[\alpha_6] - g_{22}g_{33}g_{44}g_{55}g_{66}\log[\alpha_7] \end{pmatrix}}{\begin{pmatrix} g_{32}g_{43}g_{54}g_{65}g_{76} - g_{22}g_{43}g_{54}g_{65}g_{76} + g_{22}g_{33}g_{54}g_{65}g_{76} \\ - g_{22}g_{33}g_{44}g_{65}g_{76} + g_{22}g_{33}g_{44}g_{55}g_{76} - g_{22}g_{33}g_{44}g_{55}g_{66} \end{pmatrix}};$$

$$g'_{21} = \frac{g_{21}g_{32}g_{43}g_{54}g_{65}g_{76}}{(g_{32}g_{43}g_{54}g_{65}g_{76} - g_{22}g_{43}g_{54}g_{65}g_{76} + g_{22}g_{33}g_{54}g_{65}g_{76} - g_{22}g_{33}g_{44}g_{65}g_{76} + g_{22}g_{33}g_{44}g_{55}g_{76} - g_{22}g_{33}g_{44}g_{55}g_{66})}$$

$$i = 3\text{: } \log[\alpha'_3] = \frac{(g_{43}g_{54}g_{65}g_{76}\log[\alpha_3] - g_{33}g_{54}g_{65}g_{76}\log[\alpha_4] + g_{33}g_{44}g_{65}g_{76}\log[\alpha_5] - g_{33}g_{44}g_{55}g_{76}\log[\alpha_6] + g_{33}g_{44}g_{55}g_{66}\log[\alpha_7])}{g_{43}g_{54}g_{65}g_{76} - g_{33}g_{54}g_{65}g_{76} + g_{33}g_{44}g_{65}g_{76} - g_{33}g_{44}g_{55}g_{76} + g_{33}g_{44}g_{55}g_{66}};$$

$$g'_{32} = \frac{g_{32}g_{43}g_{54}g_{65}g_{76}}{g_{43}g_{54}g_{65}g_{76} - g_{33}g_{54}g_{65}g_{76} + g_{33}g_{44}g_{65}g_{76} - g_{33}g_{44}g_{55}g_{76} + g_{33}g_{44}g_{55}g_{66}}$$

$$i = 4: \log[\alpha_4'] = \frac{g_{54}g_{65}g_{76}\log[\alpha_4] - g_{44}g_{65}g_{76}\log[\alpha_5] + g_{44}g_{55}g_{76}\log[\alpha_6] - g_{44}g_{55}g_{66}\log[\alpha_7]}{g_{54}g_{65}g_{76} - g_{44}g_{65}g_{76} + g_{44}g_{55}g_{76} - g_{44}g_{55}g_{66}};$$

$$g_{43}' = \frac{g_{43}g_{54}g_{65}g_{76}}{g_{54}g_{65}g_{76} - g_{44}g_{65}g_{76} + g_{44}g_{55}g_{76} - g_{44}g_{55}g_{66}}$$

$$i = 5: \log[\alpha_5'] = (g_{65}g_{76}\log[\alpha_5] - g_{55}g_{76}\log[\alpha_6] + g_{55}g_{66}\log[\alpha_7])/(g_{65}g_{76} - g_{55}g_{76} + g_{55}g_{66});$$

$$g_{54}' = g_{54}g_{65}g_{76}/(g_{65}g_{76} - g_{55}g_{76} + g_{55}g_{66})$$

$$i = 6: \log[\alpha_6'] = (g_{76}\log[\alpha_6] - g_{66}\log[\alpha_7])/(g_{76} - g_{66}); \quad g_{65}' = g_{65}g_{76}/(g_{76} - g_{66})$$

$n = 7$

$$i = 1: \log[\alpha_1'] = \log[\alpha_1] - g_{11}\frac{\begin{pmatrix} g_{32}g_{43}g_{54}g_{65}g_{76}\log[\alpha_2/\alpha_8] - g_{22}g_{43}g_{54}g_{65}g_{76}\log[\alpha_3/\alpha_8] + g_{22}g_{33}g_{54}g_{65}g_{76}\log[\alpha_4/\alpha_8] \\ - g_{22}g_{33}g_{44}g_{65}g_{76}\log[\alpha_5/\alpha_8] + g_{22}g_{33}g_{44}g_{55}g_{76}\log[\alpha_6/\alpha_8] - g_{22}g_{33}g_{44}g_{55}g_{66}\log[\alpha_7/\alpha_8] \end{pmatrix}}{g_{21}g_{32}g_{43}g_{54}g_{65}g_{76}};$$

$$g_{17}' = g_{17} + g_{11}\frac{(g_{32}g_{43}g_{54}g_{65}g_{76} - g_{22}g_{43}g_{54}g_{65}g_{76} + g_{22}g_{33}g_{54}g_{65}g_{76} - g_{22}g_{33}g_{44}g_{65}g_{76} + g_{22}g_{33}g_{44}g_{55}g_{76} - g_{22}g_{33}g_{44}g_{55}g_{66}}{g_{21}g_{32}g_{43}g_{54}g_{65}g_{76}}$$

$$i = 2: \log[\alpha_2'] = \frac{\begin{pmatrix} g_{32}g_{43}g_{54}g_{65}g_{76}g_{87}\log[\alpha_2] - g_{22}g_{43}g_{54}g_{65}g_{76}g_{87}\log[\alpha_3] \\ + g_{22}g_{33}g_{54}g_{65}g_{76}g_{87}\log[\alpha_4] - g_{22}g_{33}g_{44}g_{65}g_{76}g_{87}\log[\alpha_5] \\ + g_{22}g_{33}g_{44}g_{55}g_{76}g_{87}\log[\alpha_6] - g_{22}g_{33}g_{44}g_{55}g_{66}g_{87}\log[\alpha_7] + g_{22}g_{33}g_{44}g_{55}g_{66}g_{77}\log[\alpha_8] \end{pmatrix}}{\begin{pmatrix} g_{32}g_{43}g_{54}g_{65}g_{76}g_{87} - g_{22}g_{43}g_{54}g_{65}g_{76}g_{87} + g_{22}g_{33}g_{54}g_{65}g_{76}g_{87} \\ - g_{22}g_{33}g_{44}g_{65}g_{76}g_{87} + g_{22}g_{33}g_{44}g_{55}g_{76}g_{87} - g_{22}g_{33}g_{44}g_{55}g_{66}g_{87} \\ + g_{22}g_{33}g_{44}g_{55}g_{66}g_{77} \end{pmatrix}};$$

$$g_{21}' = \frac{g_{21}g_{32}g_{43}g_{54}g_{65}g_{76}g_{87}}{\begin{pmatrix} g_{32}g_{43}g_{54}g_{65}g_{76}g_{87} - g_{22}g_{43}g_{54}g_{65}g_{76}g_{87} + g_{22}g_{33}g_{54}g_{65}g_{76}g_{87} \\ - g_{22}g_{33}g_{44}g_{65}g_{76}g_{87} + g_{22}g_{33}g_{44}g_{55}g_{76}g_{87} - g_{22}g_{33}g_{44}g_{55}g_{66}g_{87} \\ + g_{22}g_{33}g_{44}g_{55}g_{66}g_{77} \end{pmatrix}}$$

$$i = 3: \log[\alpha_3'] = \frac{\begin{pmatrix} g_{43}g_{54}g_{65}g_{76}g_{87}\log[\alpha_3] - g_{33}g_{54}g_{65}g_{76}g_{87}\log[\alpha_4] \\ + g_{33}g_{44}g_{65}g_{76}g_{87}\log[\alpha_5] - g_{33}g_{44}g_{55}g_{76}g_{87}\log[\alpha_6] \\ + g_{33}g_{44}g_{55}g_{66}g_{87}\log[\alpha_7] - g_{33}g_{44}g_{55}g_{66}g_{77}\log[\alpha_8] \end{pmatrix}}{(g_{43}g_{54}g_{65}g_{76}g_{87} - g_{33}g_{54}g_{65}g_{76}g_{87} + g_{33}g_{44}g_{65}g_{76}g_{87} - g_{33}g_{44}g_{55}g_{76}g_{87} + g_{33}g_{44}g_{55}g_{66}g_{87} - g_{33}g_{44}g_{55}g_{66}g_{77})};$$

$$g_{32}' = \frac{g_{32}g_{43}g_{54}g_{65}g_{76}g_{87}}{(g_{43}g_{54}g_{65}g_{76}g_{87} - g_{33}g_{54}g_{65}g_{76}g_{87} + g_{33}g_{44}g_{65}g_{76}g_{87} - g_{33}g_{44}g_{55}g_{76}g_{87} + g_{33}g_{44}g_{55}g_{66}g_{87} - g_{33}g_{44}g_{55}g_{66}g_{77})}$$

$$i = 4: \log[\alpha_4'] = \frac{(g_{54}g_{65}g_{76}g_{87}\log[\alpha_4] - g_{44}g_{65}g_{76}g_{87}\log[\alpha_5] + g_{44}g_{55}g_{76}g_{87}\log[\alpha_6] - g_{44}g_{55}g_{66}g_{87}\log[\alpha_7] + g_{44}g_{55}g_{66}g_{77}\log[\alpha_8])}{g_{54}g_{65}g_{76}g_{87} - g_{44}g_{65}g_{76}g_{87} + g_{44}g_{55}g_{76}g_{87} - g_{44}g_{55}g_{66}g_{87} + g_{44}g_{55}g_{66}g_{77}};$$

$$g_{43}' = \frac{g_{43}g_{54}g_{65}g_{76}g_{87}}{g_{54}g_{65}g_{76}g_{87} - g_{44}g_{65}g_{76}g_{87} + g_{44}g_{55}g_{76}g_{87} - g_{44}g_{55}g_{66}g_{87} + g_{44}g_{55}g_{66}g_{77}}$$

$$i = 5: \log[\alpha_5'] = \frac{g_{65}g_{76}g_{87}\log[\alpha_5] - g_{55}g_{76}g_{87}\log[\alpha_6] + g_{55}g_{66}g_{87}\log[\alpha_7] - g_{55}g_{66}g_{77}\log[\alpha_8]}{g_{65}g_{76}g_{87} - g_{55}g_{76}g_{87} + g_{55}g_{66}g_{87} - g_{55}g_{66}g_{77}};$$

$$g_{54}' = g_{54}g_{65}g_{76}g_{87}/(g_{65}g_{76}g_{87} - g_{55}g_{76}g_{87} + g_{55}g_{66}g_{87} - g_{55}g_{66}g_{77})$$

$$i = 6: \log[\alpha_6'] = (g_{76}g_{87}\log[\alpha_6] - g_{66}g_{87}\log[\alpha_7] + g_{66}g_{77}\log[\alpha_8])/(g_{76}g_{87} - g_{66}g_{87} + g_{66}g_{77});$$

$$g_{65}' = g_{65}g_{76}g_{87}/(g_{76}g_{87} - g_{66}g_{87} + g_{66}g_{77})$$

$$i = 7: \log[\alpha_7'] = (g_{87}\log[\alpha_7] - g_{77}\log[\alpha_8])/(g_{87} - g_{77}); \quad g_{76}' = g_{76}g_{87}/(g_{87} - g_{77})$$

This work was supported in part by a joint Ph.D. fellowship PRAXIS XXI/BD/9803/96 granted by PRAXIS XXI through Programa Gulbenkian de Doutoramentos em Biologia e Medicina (to R. A.), U.S. Public Health Service grant RO1-GM30054 from the National Institutes of Health (to M. A. S.), and U.S. Department of Defense grant N00014-97-1-0364 from the Office of Naval Research (to M. A. S.). We thank Armindo Salvador for critically reading early versions of this manuscript and making useful comments.

REFERENCES

Albery, W. J., and J. R. Knowles. 1976. Evolution of enzyme function and the development of catalytic efficiency. *Biochemistry.* 15:5631–5639.

Alves, R., and M. A. Savageau. 2000a. Comparing systemic properties of ensembles of biological networks by graphical and statistical methods. *Bioinformatics.* 16:527–533.

Alves, R., and M. A. Savageau. 2000b. Systemic properties of ensembles of metabolic networks: application of graphical and statistical methods to simple unbranched pathways. *Bioinformatics.* 16:534–547.

Alves, R., and M. A. Savageau. 2000c. Extending the method of mathematically controlled comparison to include numerical comparisons. *Bioinformatics.* 16:786–798.

Alves, R., and M. A. Savageau. 2000d. Effect of overall feedback inhibition in unbranched biosynthetic pathways. *Biophys. J.* 79:2290–2304.

Atkinson, D. E. 1969. Limitation of metabolite concentrations and the conservation of solvent capacity in the living cell. *Curr. Top. Cell Reg.* 1:29–43.

Bish, D. R., and M. L. Mavrovouniotis. 1998. Enzymatic reaction rate limits with constraints on equilibrium constants and experimental parameters. *Biosystems.* 47:37–60.

Cornish-Bowden, A. 1976. The effect of natural selection on enzymic catalysis. *J. Mol. Biol.* 101:1–9.

Crowley, P. H. 1975. Natural selection and the Michaelis constant. *J. Theor. Biol.* 50:461–475.

Dorf, R. C. 1992. Modern Control Systems, 6th ed., Addison-Wesley, Reading, MA.

Fersht, A. R. 1974. Catalytic, binding and enzyme-substrate complementarity. *Proc. R. Soc.* 187:397–407.

Heinrich, R., and E. Hoffman. 1991. Kinetic parameters of enzymatic reaction states of maximal activity: an evolutionary approach. *J. Theor. Biol.* 151:249–283.

Heinrich, R., and E. Klipp. 1996. Control analysis of unbranched enzymatic chains in states of maximal activity. *J. Theor. Biol.* 182:243–252.

Heinrich, R., and S. Schuster. 1998. The modeling of metabolic systems, structure, control and optimality. *Biosystems.* 47:61–77.

Hlavacek, W. S., and M. A. Savageau. 1997. Completely uncoupled and perfectly coupled gene expression in repressible pathways. *J. Mol. Biol.* 266:538–558.

Irvine, D. H., and M. A. Savageau. 1985. Network regulation of the immune response: alternative control points for suppressor modulation of effector lymphocytes. *J. Immunol.* 134:2100–2116.

Klipp, E., and R. Heinrich. 1994. Evolutionary optimization of enzyme kinetic parameters—effect of constraints. *J. Theor. Biol.* 171:309–323.

Levine, R. L., and A. Ginsburg. 1985. Modulation by molecular interactions. *Curr. Top. Cell Reg.* 26:1–549.

Mavrovouniotis, M. L., G. Stephanopoulos, and G. Stephanopoulos. 1990. Estimation of upper bounds for the rate of enzymic reactions. *Chem. Eng. Commun.* 93:211–236.

Monod, J., J.-P. Changeux, and F. Jacob. 1963. Allosteric proteins and cellular control systems. *J. Mol. Biol.* 6:306–329.

Peterson, G. 1992. Evolutionary optimization of the catalytic efficiency of enzymes. *Eur. J. Biochem.* 206:289–295.

Peterson, G. 1996. A new approach for determination of the selectively favored kinetic design of enzyme reactions. *J. Theor. Biol.* 183:179–183.

Savageau, M. A. 1969. Biochemical systems analysis II: the steady state solution for an n-pool system using a power law approximation. *J. Theor. Biol.* 25:370–379.

Savageau, M. A. 1971a. Concepts relating the behavior of biochemical systems to their underlying molecular properties. *Arch. Biochem. Biophys.* 145:612–621.

Savageau, M. A. 1971b. Parameter sensitivity as a criterion for evaluating and comparing the performance of biochemical systems. *Nature.* 229:542–544.

Savageau, M. A. 1972. The behavior of intact biochemical control systems. *Curr. Top. Cell Reg.* 6:63–130.

Savageau, M. A. 1974. Optimal design of feedback control by inhibition: steady-state considerations. *J. Mol. Evol.* 4:139–156.

Savageau, M. A. 1975. Optimal design of feedback control by inhibition: dynamical considerations. *J. Mol. Evol.* 5:199–222.

Savageau, M. A. 1976. Biochemical Systems Analysis: A Study of Function and Design in Molecular Biology. Addison-Wesley, Reading, MA.

Savageau, M. A. 1996. A kinetic formalism for integrative molecular biology: manifestation in biochemical systems theory and use in elucidating design principles for gene circuits. *In* Integrative Approaches to Molecular Biology. J. Collado-Vides, B. Magasanik, and T. F. Smith, editors. MIT Press, Cambridge, MA. 115–146.

Savageau, M. A., and G. Jacknow. 1979. Feedforward inhibition in biosynthetic pathways: inhibition of the aminoacyl-tRNA synthetase by intermediates of the pathway. *J. Theor. Biol.* 77:405–425.

Schuster, S., and R. Heinrich. 1987. Time hierarchy in enzymatic reaction chains resulting from optimality principles. *J. Theor. Biol.* 129:189–209.

Shiraishi, F., and M. A. Savageau. 1992. The tricarboxylic acid cycle in *Dictyostelium discoideum* II. Evaluation of model consistency and robustness. *J. Biol. Chem.* 267:22919–22925.

Sorribas, A., and M. A. Savageau. 1989. Strategies for representing metabolic pathways within biochemical systems theory: reversible pathways. *Math. Biosci.* 94:239–269.

Srere, P. 1987. Complexes of sequential metabolic enzymes. *Ann. Rev. Biochem.* 56:89–124.

Umbarger, H. E. 1956. Evidence for a negative-feedback mechanism in the biosynthesis of isoleucine. *Science.* 123:848.

Waley, S. G. 1964. A note on kinetics of multienzyme systems. *Biochem. J.* 91:514–517.

Wilhelm, T., E. H. Klipp, and R. Heinrich. 1994. An evolutionary approach to enzyme kinetics: optimization of ordered mechanisms. *Bull. Math. Biol.* 56:65–106.

Wolfram, S. 1997. Mathematica™: A System for Doing Mathematics by Computer. Addison-Wesley, Menlo Park, CA.

Yates, R. A., and A. B. Pardee. 1956. Control of pyrimidine biosynthesis in *E. coli* by a feedback mechanism. *J. Biol. Chem.* 221:757–770.